STOCHASTIC VOLATILITY MODELS WITH PERSISTENT LATENT FACTORS: THEORY AND ITS APPLICATIONS TO ASSET PRICES

A Dissertation

by

HYOUNG IL LEE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2008

Major Subject: Economics

STOCHASTIC VOLATILITY MODELS WITH PERSISTENT LATENT FACTORS: THEORY AND ITS APPLICATIONS TO ASSET PRICES

A Dissertation

by

HYOUNG IL LEE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Joon Y. Park
Committee Members,	Dennis W. Jansen
	Hwagyun Kim
	Ximing Wu
Head of Department,	Larry Oliver

August 2008

Major Subject: Economics

ABSTRACT

Stochastic Volatility Models
with Persistent Latent Factors:
Theory and Its Applications to Asset Prices. (August 2008)
Hyoung Il Lee, B.A., Seoul National University
Chair of Advisory Committee: Dr. Joon Y. Park

We consider the stochastic volatility model with smooth transition and persistent latent factors. We argue that this model has advantages over the conventional stochastic model for the persistent volatility factor. Though the linear filtering is widely used in the state space model, the simulation result, as well as theory, shows that it does not work in our model. So we apply the density-based filtering method; in particular, we develop two methods to get solutions. One is the conventional approach using the Maximum Likelihood estimation and the other is the Bayesian approach using Gibbs sampling. We do a simulation study to explore their characteristics, and we apply both methods to actual macroeconomic data to extract the volatility generating process and to compare macro fundamentals with them.

Next we extend our model into multivariate model extracting common and idiosyncratic volatility for multivariate processes. We think it is interesting to apply this multivariate model into measuring time-varying uncertainty of macroeconomic variables and studying the links to market returns via a consumption-based asset pricing model. Motivated by Bansal and Yaron (2004), we extract a common volatility factor using consumption and dividend growth, and we find that this factor predicts post-war business cycle recessions quite well. Then, we estimate a long-run risk model of asset prices incorporating this macroeconomic uncertainty. We find that both risk

aversion and the intertemporal elasticity of substitution are estimated to be around two, and our simulation results show that the model can match the first and second moments of market return and risk-free rate, hence the equity premium.

To my family

ACKNOWLEDGMENTS

Needless to say, I am indebted to my advisor Joon Y. Park for his insightful guidance and firm belief in me. Each time I encountered an obstacle, he suggested methods by which I could break through. I would also like to thank Dennis Jansen, Hwagyun Kim, Ximing Wu and Issac Miller[†] for their valuable suggestions. I would like to thank especially Joon Y. Park, Hwagyun Kim and Issac Miller, who are my co-authors and who were heavily involved in writing my dissertation either at the stage of the draft or the final touches. I would like to acknowledge the Department of Economics and Program for Time Series with Applications in Financial Economics (PTSAFE), in which I have discussed interesting features of my model and built a foundation on econometrics and macroeconomics. I owe much to my colleagues Yongok Choi, Daehee Jeong, Minsoo Jeong and Hyosung Yeo. They helped me a lot in running the program and writing the paper. And thank you to Jongbum Son, who helped in making codes.

The encouragement of my parents Jung Tae Lee and Young Sook Seo towards my academic achievements was a motivating force whenever I faced difficulties. My parents in law, Kwonhee Lee and Moonja Han, also encouraged my progress and cared much about my well-being. Special thanks are reserved for my wife, Ja Young. She was always there whenever I needed her. Lastly thanks to my two lovely children, Ji Sun and Hyun Kyu, for revitalizing my life every day. I love you all.

[†]Department of Economics, University of Missouri

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

CHAPTER I

INTRODUCTION

Recently, stochastic volatility models have drawn much attention. It has long been customary to use ARCH and GARCH-type processes to model volatility in many economic and financial time series. The ARCH and GARCH-type models were quite popular and successful. However, it is now widely recognized that volatility is not entirely predictable and driven by a shock not perfectly correlated with the past values of the underlying process. The conventional volatility model specifies the conditional variance of the underlying process as an AR process in logarithm. The motivation for taking logarithm is clear: It allows us to readily impose the required restriction that the conditional variances should be positive.

In the chapter II, I consider a new stochastic volatility model possibly with the leverage effect. In our model, the stochastic volatility is assumed to be generated by a logistic transformation of an AR latent factor. The logistic function has several desirable properties to be used in the volatility model. In particular, it may be interpreted as representing the volatility levels in two regimes, i.e., high and low volatility regimes, with smooth transition between them. Indeed, there is a strong evidence for regime shifting in volatility, as shown by So, Lam and Li (1998). Our model is different from the usual regime switching model, which presumes an exogenous and abrupt change in switching regimes. I believe that smooth transition is much more realistic than discontinuous jump in modelling volatility with regime shifting.

Our stochastic volatility model with logistic volatility function and (nearly) inte-

---

This dissertation follows the style of *Econometrica*.

grated volatility factor yields many interesting time series properties such as volatility clustering and leptokurtosis. In particular, the sample autocorrelation function of the process does not decay at all and has a common distribution asymptotically independent of leads and lags, yielding a very realistic pattern of volatility clusterings. Moreover, the sample kurtosis of the process does not converge to a constant. Instead it weakly converges to a distribution with support truncated on the left by the kurtosis of the innovation. Therefore, it is expected to generated scattered numbers that are all greater than three, the kurtosis of standard normal distribution. This is quite consistent with common characteristics of many economic and financial data.

To statistically analyze our model, I consider two approaches: The conventional approach using the density-based Kalman filter, and the Bayesian approach using the Gibbs sampling. For the first approach, I numerically compute the conditional densities in the usual three steps, prediction, updating and smoothing, of the Kalman filter. The unknown parameters are estimated by the maximum likelihood (ML) method using the numerical likelihood function obtained in implementing the Kalman filter. The second approach applies the standard Bayesian procedure and estimates the unknown parameters and latent factor by their posterior means, which are computed by simulations through Gibbs sampling. The second approach is computationally more burdensome, but can be more readily extended to a model with multi-factors, compared with the first approach.

I consider two illustrative empirical examples in the paper. For both examples, our model and methodology are applied to estimate the unknown parameters and extract the latent volatility factor. The first example studies the volatility of S&P 500 returns. The results are robust and look very interesting. In particular, the extracted volatility factor seems to share a common trend with the detrended GDP series over most of the sample period. The second example examines the volatility

of dividend growth. The results appear to be quite sensible and unambiguous. The extracted volatility factor has some obvious common trend as nominal interest rates. Two of our approaches yield remarkably similar results on the estimation of parameter values and the extraction of volatility factor. I find that our model and methodology are very useful in investigating the effect of macroeconomic fundamentals on volatility of the financial market.

In the chapter III, I try to measure time-varying uncertainty of macroeconomic variables and studying the links to market returns via estimating a consumption-based asset pricing model with a non-expected utility function. Time-varying macroeconomic uncertainty is an important ingredient for asset valuation. Due to the nature of aggregate shocks, macroeconomic uncertainty is reflected in equilibrium asset prices because asset holders will demand some premium for bearing such undiversified risk. However, macroeconomic uncertainty is not an observable to economists, and therefore modelling and measuring this uncertainty is a meaningful, yet challenging task. Furthermore, the most popular macroeconomic asset pricing models identify consumption growth process as the link between macroeconomic variables and asset returns, exploiting the simple and elegant Euler equation of consumption growth and asset returns. Alas, aggregate consumption is close to a random walk and the size of unconditional variance of consumption growth is fairly modest to justify high average equity premium with low and stable interest rates. These puzzles based on consumption based asset pricing models have been one of the main research questions in finance and macroeconomics since Hansen and Singleton (1982) and Mehra and Prescott (1985). In this paper, I tackle this issue by measuring time-varying uncertainty of macroeconomic variables and studying the links to market returns via estimating a consumption-based asset pricing model with a non-expected utility function. As the first step, I develop a stochastic volatility model and propose an

econometric procedure to extract common and idiosyncratic volatility for multivariate processes. Novel features of our volatility model include a unit root common factor and logistic volatility function. This setup allows a persistent conditional volatility shifting between two regimes (high and low uncertainty regimes) with a smooth transition. The existence of transition period between high and low volatility regimes implies that economic agents may dislike transition periods because of uncertainty in regimes that they belong to. Epstein and Zin (1989) show that economic agents prefer an earlier resolution of uncertainty if risk aversion parameter is bigger than the reciprocal of elasticity of intertemporal substitution in case of Kreps-Porteus utility function. Thus, together with this non-expected utility function, our volatility setup can generate a higher risk premium even with medium level of volatility.

In terms of econometric setup, our volatility model can be regarded as a non-linear, non-stationary state space model. There are several non-linear filtering techniques in the conventional approach to solve such a stochastic volatility model with nonlinear measurement equation. But multi-dimensionality of our problem makes these filtering techniques much more difficult to be applied. To overcome this issue, I use Gibbs sampling which does not suffer seriously from the curse of dimensionality, because it utilizes univariate conditional density function. Here I develop an algorithm to filter macroeconomic uncertainty based on the chapter II which studies the univariate stochastic volatility model with a logistic function.

When I put our methods into data, since I am interested in performances of consumption based asset pricing model, I extract a common and idiosyncratic volatility factors for consumption and dividend growth. I find that the common factor delineating 'macroeconomic uncertainty' predicts post-war business cycle recessions quite well. Then, I estimate a long-run risk model of asset prices incorporating this macroeconomic uncertainty. According to our estimation, both risk aversion and the

intertemporal elasticity of substitution are estimated around two, and our simulation results show that the model can match the first and second moments of market return and risk-free rate, hence explains the equity premium, the risk-free rate puzzle, and volatility puzzle.

In the chapter IV, I try to compare the performance of the Extended Kalman Filter (EKF) with the Density-based Nonlinear Kalman Filter (DNF). When the functional form in the measurement equation is logistic, the EKF is more difficult to be applied than the DNF because of integrability of the derivative of logistic function. In other words, Kalman gain disappear as volatility factor goes to high or low volatility regime. I show this fact by the extensive simulation. First I try to Monte-Carlo experiments with stationary latent volatility factor and nonstationary latent volatility factor. I see that nonstationarity makes the EKF estimation poorer while the DNF has relatively similar results. Secondly, I change the functional form into power function. I see the DNF's performance better than the EKF.

CHAPTER II

STOCHASTIC VOLATILITY MODELS WITH SMOOTH TRANSITION REGIMES

A. Introduction

Recently, stochastic volatility models have drawn much attention. It has long been customary to use ARCH and GARCH-type processes to model volatility in many economic and financial time series. The ARCH and GARCH-type models were quite popular and successful. However, it is now widely recognized that volatility is not entirely predictable and driven by a shock not perfectly correlated with the past values of the underlying process. The conventional volatility model specifies the conditional variance of the underlying process as an AR process in logarithm. The motivation for taking logarithm is clear: It allows us to readily impose the required restriction that the conditional variances should be positive. Several authors including Jacquier, Polson, and Rossi (1994) use this model to investigate the stochastic volatility of stock returns.

In the paper, I consider a new stochastic volatility model possibly with the leverage effect. In our model, the stochastic volatility is assumed to be generated by a logistic transformation of an AR latent factor. The logistic function has several desirable properties to be used in the volatility model. In particular, it may be interpreted as representing the volatility levels in two regimes, i.e., high and low volatility regimes, with smooth transition between them. Indeed, there is a strong evidence for regime shifting in volatility, as shown by So, Lam and Li (1998). Our model is different from the usual regime switching model, which presumes an exogenous and abrupt change in switching regimes. I believe that smooth transition is much more

realistic than discontinuous jump in modelling volatility with regime shifting.

For all economic and financial time series I investigated, the extracted volatility factor has the AR coefficient that is extremely close to unity. For practical relevancy, it seems evident that I need to consider our model with the volatility factor having a root in a close proximity of the unity. In fact, the exact or near integratedness of volatility process has been observed widely in many different contexts. In predominantly many cases, ARCH and GARCH-type models are estimated as being very close to integrated ARCH and GARCH-type models. The same is true in stochastic volatility models. For instance, the estimate for the AR coefficient of the log volatility process that Jacquier, Polson, and Rossi (1994), Jacquier, Polson, and Rossi (2004) obtained for stock returns was around .95.

Our stochastic volatility model with logistic volatility function and (nearly) integrated volatility factor yields many interesting time series properties. The reader is referred to Park (2002) for the details. Indeed, the process generated by our stochastic volatility model has many characteristics that are consistent with what I typically observe from many economic financial data such as volatility clustering and leptokurtosis. In particular, the sample autocorrelation function of the process does not decay at all and has a common distribution asymptotically independent of leads and lags, yielding a very realistic pattern of volatility clusterings. Moreover, the sample kurtosis of the process does not converge to a constant. Instead it weakly converges to a distribution with support truncated on the left by the kurtosis of the innovation. Therefore, it is expected to generated scattered numbers that are all greater than three, the kurtosis of standard normal distribution. This is quite consistent with common characteristics of many economic and financial data.

The stochastic volatility model considered previously by Jacquier, Polson, and Rossi (1994), Jacquier, Polson, and Rossi (2004) and So, Lam, and Li (1998) may be

viewed as the same model as ours with an exponential volatility function, in place of a logistic volatility function. The model has several undesirable properties, especially given the strong empirical evidence for the (near) nonstationarity of volatility factor. The exponential function is unbounded and too explosive. In the presence of nonstationarity in volatility factor, the stochastic volatility model with an exponential volatility function generates unrealistically exploding samples as the sample size increases. Also, the model necessarily implies that the effect of a shock to the volatility factor is amplified at an exponential rate as the level of volatility increases. Furthermore, a positive shock to the volatility factor, both direct and indirect through the leverage effect, always results in a larger effect on volatility level than a negative shock.

To statistically analyze our model, I consider two approaches: The conventional approach using the density-based Kalman filter, and the Bayesian approach using the Gibbs sampling. For the first approach, I numerically compute the conditional densities in the usual three steps, prediction, updating and smoothing, of the Kalman filter. The unknown parameters are estimated by the maximum likelihood (ML) method using the numerical likelihood function obtained in implementing the Kalman filter. See, e.g., Tanizaki (1996) for a general introduction to the density-based Kalman filter. The second approach applies the standard Bayesian procedure and estimates the unknown parameters and latent factor by their posterior means, which are computed by simulations through Gibbs sampling. Our Bayesian methodology is largely the same as the one adopted by Jacquier, Polson, and Rossi (1994) and Jacquier, Polson, and Rossi (2004). The second approach is computationally more burdensome, but can be more readily extended to a model with multi-factors, compared with the first approach.

I consider two illustrative empirical examples in the paper. For both examples,

our model and methodology are applied to estimate the unknown parameters and extract the latent volatility factor. The first example studies the volatility of S&P 500 returns. The results are robust and look very interesting. In particular, the extracted volatility factor seems to share a common trend with the detrended GDP series over most of the sample period. The second example examines the volatility of dividend growth. The results appear to be quite sensible and unambiguous. The extracted volatility factor has some obvious common trend as nominal interest rates. Two of our approaches yield remarkably similar results on the estimation of parameter values and the extraction of volatility factor. I find that our model and methodology are very useful in investigating the effect of macroeconomic fundamentals on volatility of the financial market.

The rest of the paper is organized as follows. In Section B, I introduce our stochastic volatility model, and compares it with other existing models. In particular, I parametrically specify the volatility function as a logistic function and introduce the leverage effect. Then I compare our model with the existing model using an exponential volatility function, especially by probing what distinctive characteristics these two models have in contrast to each other. Sections C and D respectively introduce the conventional and Bayesian approaches to statistically analyze our model. Their algorithms are explained in detail, and some simulation results are presented to show their performances. Empirical applications are given in Section E. I provide the empirical results for two illustrative examples, which deal with the stock return and dividend volatilities respectively. In particular, the extracted volatility factors are tested to see whether they have common stochastic trends with other observable macroeconomic variables such as detrended DGP and nominal interest rates. Section F concludes the paper.

B. The Model

In this section, I introduce our stochastic volatility model and make comparisons with other existing models.

1. A Stochastic Volatility Model

I consider the nonlinear nonstationary state-space model given by

$$y_t = \sqrt{f(x_t, \beta)}\, u_t,$$
$$x_{t+1} = \alpha\, x_t + v_{t+1}, \tag{2.1}$$

where I make the following assumptions:

Assumption 1: The volatility function is given by

$$f(x_t, \beta) = \mu + \frac{\nu}{1 + \exp(-\lambda(x_t - \kappa))}, \tag{2.2}$$

where $\beta = (\mu, \nu, \lambda, \kappa)\prime$ is a vector of unknown parameters satisfying $\mu > 0, \nu > 0$, and $\lambda > 0$.

When I model the volatility, it is plausible to have a lowerbound because the volatility cannot take negative value and upperbound because explosive volatility has not been observed given the historical evidence. The parameters μ and $\mu + \nu$ represent two asymptotic levels, i.e., low volatility regime and high volatility regime, respectively. The assumption of positivity of ν makes the logistic function upward sloping and hence, larger volatility factor means higher volatility. This is not a restriction but for the convenience because the latent variable will be extracted reversely when logistic function is downward sloping. Also, since sudden change from one regime into the other regime is not realistic I need a smooth transition period. The parameters λ and

κ specify the transition between two regimes, i.e., the speed and the reflection point of the transition, respectively.[1]

Park (2002) shows that the model with asymptotically homogeneous functions of a integrated process has several nice statistical properties. First, the sample autocorrelations of the squared processes have the same random limit for all lags i.e., strong persistence. Secondly, the sample kurtosis has supports truncated on the left by the kurtosis of the innovations i.e., leptokurtosis. Since the logistic function belongs to the class of asymptotically homogeneous function, our model can capture the volatility clustering and fat-tail features of financial time series.

Assumption 2: (x_t) is a scalar latent volatility factor and $|\alpha| \leq 1$,

I describe the volatility factor, (x_t) explicitly in the transition equation because I am interested in the linkage between it and macro economic fundamentals. I assume that this volatility factor is scalar so that this single factor drives the fluctuation of stock return. The volatility factor could be either a macroeconomic fundamentals or a mixture of some macroeconomic fundamentals. Though it is more likely and more interesting to have unit-root or at least near-unit latent factor given the characteristics of financial time series, I do not impose an restriction directly on α. Once I find $\alpha \simeq 1$, the process can be considered to have persistence so that it can generate highly autocorrelated volatility or volatility clustering, which means high volatility is followed by another high volatility, or the other way around. This model has some advantages that I can directly compare them with a persistent time series, as which many interesting macro time series can be characterized, without any taking difference of both series.

---

[1] I may think X $= [\kappa - \frac{\log(2+\sqrt{(3)})}{\lambda},\ \kappa - \frac{\log(2-\sqrt{(3)})}{\lambda}]$ as transition period because $\frac{\partial^3 f(x)}{\partial x^3} = 0$ at $x = \kappa - \frac{\log(2\pm\sqrt{(3)})}{\lambda}$.

Assumption 3: (y_t) is also a scalar variable but observable,

If I choose high frequency data as (y_t), then I can make lower frequency volatility factor with simple calculation. For example, when I use weekly data as (y_t) I can extract a daily volatility factor and construct a monthly or quarterly volatility factor. Then I can easily compare it with many macro time series. Here I do not introduce multivariate of (y_t), though I may consider it as more interesting setting. Multivariate model can be used to extract common stochastic volatility factor but I need to follow less expensive solution in terms of computation because I am going to suffer the curse of dimension. Taking this point into account, however, I will develop two algorithms for this univariate model to give future research more flexibility.

Assumption 4: the error terms follows bivariate normal distribution with correlation,

$$\begin{pmatrix} u_t \\ v_{t+1} \end{pmatrix} \sim iid\ \mathbb{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \tag{2.3}$$

The correlation parameter, ρ generates leverage effect if $-1 \leq \rho < 0$. Black (1976) has first documented leverage effect, which is a decrease in stock price or stock return increases the debt-equity ratio of the firm, increases the risk of the stock and hence increase the volatility of stock return. There are two ways of models with the leverage effect. One is contemporary correlation (See Jacquier, Polson, and Rossi (2004)) and the other is inter-temporal correlation (See Harvey and Shepard (1996)). I follows latter because it maintain Martingale difference property as argued in Harvey and Shepard (1996) and Yu (2005) argues that it is empirically superior to the former when S&P500 data is used, as in this paper. The variance of (u_t) and (v_t) is set to be unity for the identification of parameters. For example, if I assume the variance

of (u_t) to be multiplied by k^2, then level parameters are to be divided by k^2.

Assumption 5: x_0 is independent of (u_t) and (v_t).

Assumptions 4 and 5 are standard and routinely imposed in this type of model.

2. Comparisons with the Existing Models

Now I compare the difference between our model and other exiting models. Let's consider a stochastic volatility model given by

$$y_t = \sqrt{h_t}\, u_t,$$
$$\log h_{t+1} = \gamma + \alpha \log h_t + w_{t+1}, \tag{2.4}$$

where

$$\begin{pmatrix} u_t \\ w_{t+1} \end{pmatrix} \sim iid\ \mathbb{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_v \\ \rho\sigma_v & \sigma_v^2 \end{pmatrix} \right),$$

This model can be transformed into the following:

$$y_t = \sqrt{\nu \exp(\lambda x_t)}\, u_t,$$
$$x_{t+1} = \alpha\, x_t + v_{t+1}, \tag{2.5}$$

where

$$\begin{pmatrix} u_t \\ v_{t+1} \end{pmatrix} \sim iid\ \mathbb{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),$$

and $h_t = \nu \exp(\lambda x_t)$, $v_t = w_t/\sigma_v$, $\nu = \exp(\gamma/(1-\alpha))^2$ and $\lambda = \sigma_v$.

Models (2.1) and (2.5) are different only up to the functional form of the volatility function. Table 1 shows the similarity of estimation results between (2.4) and

[2]If $\alpha = 1$, then $\gamma = 0$ and $\nu = 1$.

Table 1.: Estimation Result with the Conventional Volatility Models.

Models	$\hat{\alpha}$	$\hat{\sigma_v}$	$\hat{\rho}$	Data
JPR94	0.950	0.230	-	CRSP weekly ('62∼'91)
SLL98a	0.963	0.212	-	S&P500 weekly ('61∼'87)
SLL98b	0.472	0.272	-	S&P500 weekly ('61∼'87)
JPR04	0.943	0.270	-0.460	CRSP weekly ('62∼'91)
Transformed	0.959	0.194	-0.522	S&P500 weekly ('70∼'07)

Notes: JPR94 and SLL98a estimate conventional Stochastic Volatility model and SLL98b estimates conventional model with regime switching and JPR04 estimates conventional model with leverage effect.

(2.5). But I argue this exponential form of volatility function does not match well the realities. First, the conventional stochastic volatility model generates explosive volatility eventually as the number of observation increases if the data is believed to be generated from the conventional stochastic volatility model with persistent latent factor which follows either unit-root or near unit-root process. In particular, when latent factors is integrated process λ should go to zero as the number of observations goes to infinity because exponential function amplify the nonstationarity of latent factors into volatility. It seems that the conventional stochastic volatility models do not fit the nonstationary latent factor well and in our application I find an evidence that the conventional stochastic volatility models do not perform well with nonstationary latent factor.

The regime switching volatility models have come up to make up the weakness that the conventional volatility models do not have a distinction between regimes. The regime-switching models have been developed by So, Lam, and Li (1998) to connect the volatility with state of the economy by regime. It is very natural to assume that there are multiple regimes in the economy as volatility plots reveal. The main drawback of regime-switching stochastic volatility model is that the change among the states of the economy happens abruptly though transition probability from one regime

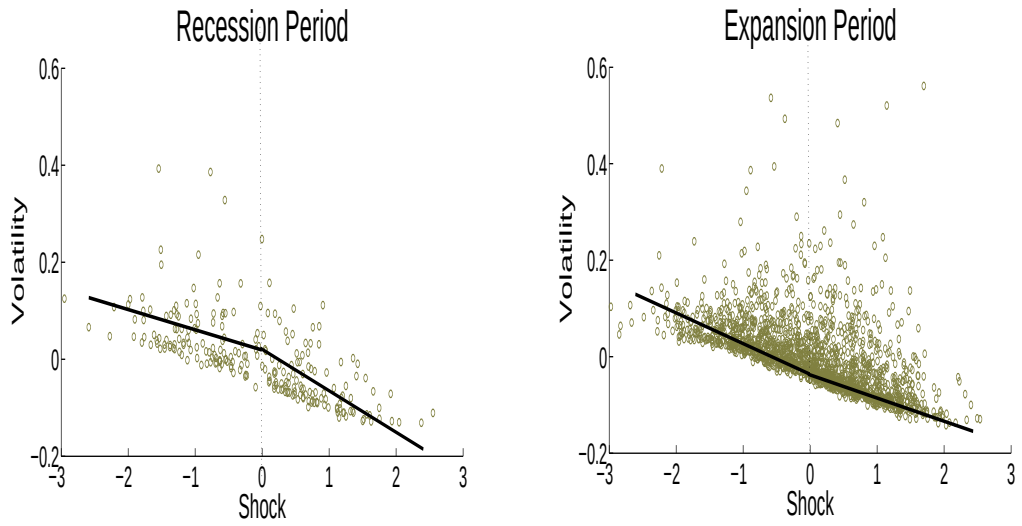

Fig. 1.: Scatter Plot of the Estimated Volatility Responses and Shocks.

to the another regime is defined. And since they estimate the parameters separately across the regime, it has more parameters to be estimated so that estimation becomes less efficient. Moreover they have still unbounded and explosive volatility in high volatility regime.

Secondly, the implication on the leverage effect is different in terms of the absolute magnitude. The conventional stochastic volatility model implies negative shock

Table 2.: Leverage Effect Quantification Examples.

	low volatility case ($x_t = -5$)		transition case ($x_t = 0$)		high volatility case ($x_t = 5$)	
	$u_t = -2$	$u_t = 2$	$u_t = -2$	$u_t = 2$	$u_t = -2$	$u_t = 2$
growth rate of the volatility	25.81	-15.33	16.24	-16.24	3.31	-8.67

Notes: Smooth transition regime volatility model with $\mu = 0.0001$, $\nu = 0.0011$, $\lambda = 0.4$,$\alpha = 1$ and $\rho = -0.5$ is used. The parameters are the same as the estimated value for S&P500 return in the section E.

has bigger impact on the volatility than positive shock, though the their directions are different, regardless of the current state i.e., high volatility or low volatility period because the slope of first derivative of the exponential function is always positive. However, the stochastic volatility models with smooth transition regimes implies asymmetric leverage effect i.e., negative shock to stock return has bigger impact on the volatility more than positive shock to stock return during low volatility regime while positive shock has bigger impact than negative shock to stock return during high volatility regime because the slope of first derivative of the logistic function changes the sign. Table 2 quantifies the magnitude of the leverage effect across the state of the economy. When the economy enters low volatility regime due to successive positive shock, the marginal effect of another positive shock is likely to be smaller than that of new negative shock. Likewise when the economy enters high volatility regime due to successive negative shock, things goes the other way around. While another negative shock becomes chronic, a new positive shock gives the economy a evidence of regime changing so that volatility decrease relatively big. Figure 1 is a scatter plot of estimated growth rate of the volatility of stock return, $\left(\hat{E}[f(x_{t+1})|\mathcal{F}_{t+1}] - \hat{E}[f(x_t)|\mathcal{F}_t]\right)/\hat{E}[f(x_t)|\mathcal{F}_t]$ and estimated shocks, $\hat{u_t}$ under *the conventional stochastic volatility models* across business cycle.[3] It shows the leverage effect and its asymmetry across the state of the economy. Left panel tells that relative importance of the positive shock to stock return during the recession while right panel tells that relative importance of the negative shock to stock return during the expansion. Note that So, Lam, and Li (1998) argues that high volatility period of stock return seems strongly associated with recession and Hamilton and Lin (1996) concludes that economic recessions are the primary factor that drives fluctuations in

[3]Stochastic volatility models with smooth transition regimes have similar scatter plot.

the volatility of stock returns.

Thirdly, the conventional stochastic volatility models are too sensitive to outliers in the sense that they estimate the volatility high when they observe high squared return, as argued in Jacquier, Polson, and Rossi (2004). But the stochastic volatility models with smooth transition regimes resist to outliers because they cannot estimate the volatility more than $\mu + \nu$ even if they observe outliers. Hence the conventional stochastic volatility models will result in a more variable sequence of estimated volatility's than smooth transition volatility models. The sensitivity to large squared return could make a difference in terms of forecasting performance because forecasting future volatility is definitely related to current volatility estimation. If current shock is not related to fundamental, then forecasting errors from the conventional stochastic volatility models become larger than the stochastic volatility models with smooth transition regimes. While Jacquier, Polson, and Rossi (2004) shows that fat-tailed model of the conventional stochastic volatility models can improve the resistance to outliers well, they need another sequence of random variable and parameter.

C. Conventional Approach Using Density-Based Filter

If I have linear state-space model, then Kalman filtering is the best and unique algorithm. For a nonlinear state-space model, however, there are several algorithms, for example, Extended Kalman filtering, Unscented Kalman filtering, and Numerical Integration filtering, Monte-carlo filtering, Gaussian Sum filtering, Importance Sampling filtering, Rejection Sampling filtering, Gibbs sampling etc. Extended Kalman filtering and Unscented Kalman filtering are based on an approximation of either nonlinear function to linear function or non Gaussian conditional density to Gaussian conditional density. The approximation to linear function is not appropriate because

the volatility function is close to flat at the extreme value which means that new information cannot feed back to previous prediction and hence, latent factor may be stuck to some high or low value. The approximation to density to Gaussian is not appropriate because the error terms is logarithm of $\chi^2(1)$ distribution which has fat tail to the left, if I take squares and logs. All the remains are density-based filters. Except for Gibbs sampling, density-based filters are implemented by maximum likelihood estimation(ML estimation). I develop ML estimation algorithm in this section and Gibbs sampling in the next section.

1. Algorithm

The density-based nonlinear filtering algorithm is to deal with whole distribution information instead of the first and second moments because I do not assume the normality of the conditional density function. Other than that the basic scheme such as the predicting, updating and smoothing steps is the same as the linear filtering. Harvey (1990) and Tanizaki (1996) show that I can construct the predicting, updating and smoothing steps using density function conditional on $\mathcal{F}_{t-1}$ or $\mathcal{F}_t$ under the assumption of independence between measurement errors and transition errors. The dependency between them in our model, however, makes the algorithm more complicated.

For the prediction step, I utilize the relationship

$$\begin{aligned}
p(x_t|\mathcal{F}_{t-1},\theta) &= \int p(x_t, x_{t-1}|\mathcal{F}_{t-1},\theta)\, dx_{t-1} \\
&= \int p(x_t|x_{t-1},\mathcal{F}_{t-1},\theta)p(x_{t-1}|\mathcal{F}_{t-1},\theta)\, dx_{t-1} \\
&= \int p(x_t|x_{t-1},y_{t-1},\mathcal{F}_{t-2},\theta)p(x_{t-1}|\mathcal{F}_{t-1},\theta)\, dx_{t-1} \\
&= \int p(x_t|x_{t-1},y_{t-1},\theta)p(x_{t-1}|\mathcal{F}_{t-1},\theta)\, dx_{t-1},
\end{aligned}$$

where $x_t|x_{t-1}, y_{t-1}, \theta \sim \mathbb{N}(\alpha x_{t-1} + \rho y_{t-1}/f(x_{t-1}, \theta), 1-\rho^2)$. Note that the last equality comes from the fact that (a) I can transform v_t into $\rho y_{t-1}/f(x_{t-1}) + \sqrt{1-\rho^2}\epsilon_t$ where ϵ_t is white noise and (b) ϵ_t and $\mathcal{F}_{t-2}$ are independent.

On the other hand, the updating step relies on the relationship

$$\begin{aligned} p(x_t|\mathcal{F}_t, \theta) &= p(x_t|y_t, \mathcal{F}_{t-1}, \theta) \\ &= \frac{p(x_t, y_t|\mathcal{F}_{t-1}, \theta)}{p(y_t|\mathcal{F}_{t-1}, \theta)} \\ &= \frac{p(x_t, y_t|\mathcal{F}_{t-1}, \theta)}{\int p(x_t, y_t|\mathcal{F}_{t-1}, \theta)\, dx_t} \\ &= \frac{p(y_t|x_t, \mathcal{F}_{t-1}, \theta) p(x_t|\mathcal{F}_{t-1}, \theta)}{\int p(y_t|x_t, \mathcal{F}_{t-1}, \theta) p(x_t|\mathcal{F}_{t-1}, \theta)\, dx_t} \\ &= \frac{p(y_t|x_t, \theta) p(x_t|\mathcal{F}_{t-1}, \theta)}{\int p(y_t|x_t, \theta) p(x_t|\mathcal{F}_{t-1}, \theta)\, dx_t}, \end{aligned}$$

where $y_t|x_t, \theta \sim \mathbb{N}(0, f(x_t, \beta))$. Note that the last equality comes from the fact that u_t and $\mathcal{F}_{t-1}$ are independent.

Since log likelihood function is given by

$$\ell(y_n, \ldots, y_1|\theta) = \sum \log p(y_t|\mathcal{F}_{t-1}, \theta),$$

ML estimator can be defined by

$$\begin{aligned} \hat{\theta} &= arg \max_{\theta \in \Theta_0} \ell(y_n, \ldots, y_1|\theta) \\ &= arg \max_{\theta \in \Theta_0} \sum_{t=1}^{n} \log p(y_t|\mathcal{F}_{t-1}, \theta), \end{aligned} \tag{2.6}$$

where $p(y_t|\mathcal{F}_{t-1}, \theta) = \int p(y_t|x_t, \theta)\ p(x_t|\mathcal{F}_{t-1}, \theta)\, dx_t$ which is available from the updating step.

In general, the conditional densities here cannot be obtained analytically. The only thing I can do is to get a numerical solution for all the conditional densities. Tanizaki (1996) suggests several methods to reduce the burden of computation, for

example, Numerical Integration filtering, Monte-Carlo filtering, Gaussian Sum filtering, Importance Sampling filtering, Rejection Sampling filtering. Among them I think the Numerical Integration filtering is the most straightforward and direct method.

But there is a big difference between Numerical Integration Filter of Tanizaki (1996) and ours in terms of choice of nodes. (I need nodes to make an rectangular for integration) For the choice of nodes Tanizaki (1996) uses random draws from the constructed intervals by the mean and variance of the latent factors based on Extended Kalman filtering. But what if Extended Kalman filtering works bad? In our model Extended Kalman filtering has bigger possibility of being stuck with higher or lower values of latent factor due to the integrability of logistic function. Once it become stuck, so is the intervals. I choose the simple equally-divided nodes for the fixed length of interval. Since the latent factors of unit-root or near unit-root process can go anywhere, however, the location of the interval should adjust depending the updated expected values of the latent factor for each time.

Now suppose that $p(x_{t-1}|\mathcal{F}_{t-1})$ is measured over $[-c + x_{t-1|t-2},\ \ c + x_{t-1|t-2}]$ and has mean $x_{t-1|t-1}$.[4] For the prediction step, using the change in variables, i.e.,

[4]In this and next paragraph θ is suppressed.

$x_{t-1} = z + x_{t-1|t-2}$

$$
\begin{aligned}
p(x_t|\mathcal{F}_{t-1}) &= \int p(x_t|x_{t-1}, y_{t-1}) p(x_{t-1}|\mathcal{F}_{t-1})\, dx_{t-1} \\
&\approx \int_{-c+x_{t-1|t-2}}^{c+x_{t-1|t-2}} p(x_t|x_{t-1}, y_{t-1}) p(x_{t-1}|\mathcal{F}_{t-1})\, dx_{t-1} \\
&= \int_{-c}^{c} p(x_t|z + x_{t-1|t-2}, y_{t-1}) p(z + x_{t-1|t-2}|\mathcal{F}_{t-1})\, dz \\
&\approx \frac{h}{\sqrt{2\pi(1-\rho^2)}} \\
&\qquad \sum_{j=1}^{m} \exp\left(- \frac{\left[x_t - (\alpha(z_j + x_{t-1|t-2}) + \rho y_{t-1} / \sqrt{f(z_j + x_{t-1|t-2})}) \right]^2}{2(1-\rho^2)} \right) \\
&\qquad p(z_j + x_{t-1|t-2}|\mathcal{F}_{t-1}),
\end{aligned}
$$

where c is the one-side length of interval for the conditional density function to be measured, h is the length of each partition in the interval, $j = 1, 2, \cdots, m$, $m = 2c/h$ so that $z = [-c, -c + h, \cdots, c - h]$.

To use the change in variables again for $x_t = z + x_{t|t-1}$, I need to know the value of $x_{t|t-1} = \alpha x_{t-1|t-1} + \int v_t p(v_t|\mathcal{F}_{t-1})\, dv_t$. The second term can be computed by using following relationship

$$
\begin{aligned}
p(v_t|\mathcal{F}_{t-1}) &= \int p(v_t|x_{t-1}, y_{t-1}) p(x_{t-1}|\mathcal{F}_{t-1})\, dx_{t-1} \\
&\approx \frac{h}{\sqrt{2\pi(1-\rho^2)}} \sum_{j=1}^{m} \exp\left(- \frac{\left[v_t - \rho y_{t-1} / \sqrt{f(z_j + x_{t-1|t-2})} \right]^2}{2(1-\rho^2)} \right) \\
&\qquad p(z_j + x_{t-1|t-2}|\mathcal{F}_{t-1}),
\end{aligned}
$$

where v_t lies in [-3,3].

Therefore, the relation $x_t = z + x_{t|t-1}$ yields

$$p(z_i + x_{t|t-1}|\mathcal{F}_{t-1}) \approx \frac{h}{\sqrt{2\pi(1-\rho^2)}} \sum_{j=1}^{m} \exp\left(-\frac{\left[z_i + x_{t|t-1} - (\alpha(z_j + x_{t-1|t-2}) + \frac{\rho y_{t-1}}{\sqrt{f(z_j + x_{t-1|t-2})}}) \right]^2}{2(1-\rho^2)} \right) p(z_j + x_{t-1|t-2}|\mathcal{F}_{t-1})$$

Note that (a) $p(z + x_{t-1|t-2}|\mathcal{F}_{t-1})$ is the density shifted from $p(x_{t-1}|\mathcal{F}_{t-1})$ by $x_{t-1|t-2}$ (b) $p(x_t|\mathcal{F}_{t-1})$ is measured over $[-c + x_{t|t-1},\ c + x_{t|t-1}]$ and has mean $x_{t|t-1}$.

For the updating step, using the change in variables, i.e., $x_t = z + x_{t|t-1}$

$$\begin{aligned} p(y_t|\mathcal{F}_{t-1}) &= \int p(y_t|x_t) p(x_t|\mathcal{F}_{t-1}, \theta)\, dx_t \\ &\approx \int_{-c+x_{t|t-1}}^{c+x_{t|t-1}} p(y_t|x_t) p(x_t|\mathcal{F}_{t-1})\, dx_t \\ &= \int_{-c}^{c} p(y_t|z + x_{t|t-1}) p(z + x_{t|t-1}|\mathcal{F}_{t-1})\, dz \\ &\approx h \sum_{j=1}^{m} \frac{1}{\sqrt{2\pi f(z_j + x_{t|t-1})}} \exp\left(-\frac{y_t^2}{2f(z_j + x_{t|t-1})} \right) p(z_j + x_{t|t-1}|\mathcal{F}_{t-1}) \end{aligned}$$

$$p(z_i + x_{t|t-1}|\mathcal{F}_t) \approx \frac{1}{\sqrt{2\pi f(z_i + x_{t|t-1})}} \exp\left(-\frac{y_t^2}{2f(z_i + x_{t|t-1})} \right) \frac{p(z_i + x_{t|t-1}|\mathcal{F}_{t-1})}{p(y_t|\mathcal{F}_{t-1})}$$

Note that (a) $p(z + x_{t|t-1}|\mathcal{F}_{t-1})$ is the density shifted from $p(x_t|\mathcal{F}_{t-1})$ by $x_{t|t-1}$ (b) $p(x_t|\mathcal{F}_t)$ is measured over $[-c + x_{t|t-1},\ c + x_{t|t-1}]$ and has mean $x_{t|t}$.

During the implementation of flexible interval method, the choice of c and h could be important. Since the bigger c or the smaller h, the more burden of computation, I need to choose optimally. When I choose c, bigger c does not make better result once c is big enough to cover whole conditional densities. The rule of thumb is that I can try several initial values and get the series of conditional densities and check whether the conditional densities are truncated before converging to zero. Unless they die

out before latent factors reach to $\pm c$, then I should choose bigger c. Regarding h, the finer h would give better result, but the marginal gain is not always bigger than marginal pain, which is the elapsed time that I will spend. So, the rule of thumb is that I choose big h at the beginning and smaller and smaller h until the change in the result is negligible. In our case, with the reasonable h and c, it takes 10 minutes usually with 2000 observations.

2. Monte-Carlo Experiments

I design the Monte-Carlo experiments as following: I use a parameter set $\beta_0 = (1, 3, 0.3, 5)$ which is close to S&P500 with normalized variance and $\alpha_0 = 1$, $\rho_0 = -0.5$ because I am interested in nonstationarity and leverage effect. I run 1000 iterations with different realizations of random variables (u_t), (v_{t+1}). Initial guess is $(2, 2, 0.5, 0, 0.7, -0.2)$. I compare the number of data, n = 1000 and n = 2000 to see whether the densities get concentrated to true value as the number of data increases. Since estimating κ given x_0 is equal to estimating x_0 given κ due to the unit-root process, I estimates x_0 with $\kappa = 0$ because I want to avoid a possible interaction between λ and κ.

As illustrated in Figure 2, as the number of data increases, all the estimates become more concentrated to true value and the biasedness disappears for some estimates. Generally the estimates for $\mu, \lambda, \alpha, \rho$ show faster convergence than those for ν, x_0.

D. Bayesian Approach Using Gibbs Sampling

Since I keep in mind a multivariate model in which ML estimation methods has a difficulty due to multiple dimension, I want to develop another approach. Here I

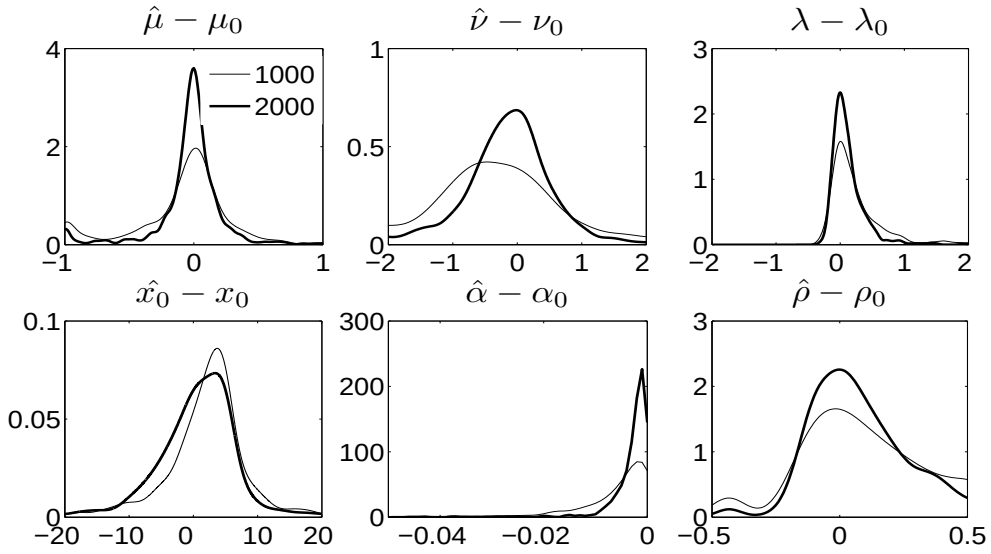

Fig. 2.: Densities of the Estimates with Monte-Carlo Experiments.

suggest a Bayesian approach i.e., Gibbs Sampling to solve this model. While Gibbs Sampling method was originally introduced for image restoration by Geman and Geman (1984), it has widely been used to solve state-space models, in particular, SV models. See the Jacquier, Polson, and Rossi (1994), So, Lam, and Li (1998), Jacquier, Polson, and Rossi (2004), and Geweke and Tanizaki (2001).

Basically Gibbs Sampling method view the estimation as sampling the parameters from the posterior density function, considering them as random numbers. I can consider sampling from $p(X, \theta|Y)$ instead of $p(\theta|Y)$ where Y is observable data and X is latent factors. This augmentation of latent factor has originally been introduced by the Tanner and Wong (1987) to calculate posterior density function, $p(\theta|Y) = \int p(X, \theta|Y)\, dX$. Once I sample (X, θ) sequentially from univariate density functions conditional on all the other information, then I have $p(X|Y) = \int p(X, \theta|Y)$ as well as $p(\theta|Y)$. The beauty of Gibbs sampling is that it is asymptotically identical

to the sampling directly from joint density functions under mild conditions[5] as proved by Tierney (1994).

1. Algorithm

I derive joint distribution of X and Y as following:

$$\begin{aligned} p(X,Y|\theta) &= p(y_n|x_n,\theta)\left(\prod_{t=1}^{n-1} p(y_t,x_{t+1}|x_t,\theta)\right)p(x_1|\theta) \\ &\propto \left(\prod_{t=1}^{n} f(x_t,\beta)^{-\frac{1}{2}}\right)(1-\rho^2)^{-\frac{n-1}{2}} \\ &\quad \exp\left[-\frac{\sum_{t=1}^{n-1}(u_t^2-2\rho u_t v_{t+1}+v_{t+1}^2)}{2(1-\rho^2)}\right]\exp\left[-\frac{1}{2}(u_n^2+v_1^2)\right] \end{aligned}$$

Then I can easily find joint posterior distribution of X and θ as following:

$$\begin{aligned} p(X,\theta|Y) &\propto \left(\prod_{t=1}^{n} f(x_t,\beta)^{-\frac{1}{2}}\right)(1-\rho^2)^{-\frac{n-1}{2}} \\ &\quad \exp\left[-\frac{\sum_{t=1}^{n-1}(u_t^2-2\rho u_t v_{t+1}+v_{t+1}^2)}{2(1-\rho^2)}\right]\exp\left[-\frac{1}{2}(u_n^2+v_1^2)\right] \\ &\quad p(x_0)p(\alpha)p(\beta)p(\rho) \end{aligned} \tag{2.7}$$

First, I derive the posterior distribution of latent volatility factor conditional on all the other information.[6] It follows readily from (3.14) that

$$\begin{aligned} p(x_t|X_{\setminus t},Y,\theta) &\propto f(x_t,\beta)^{-\frac{1}{2}}\exp\left[-\frac{u_t^2-2\rho u_t v_{t+1}}{2(1-\rho^2)}\right] \\ &\quad \exp\left[-\frac{\alpha^2+1}{2(1-\rho^2)}\left(x_t-\frac{\rho u_{t-1}+\alpha(x_{t-1}+x_{t+1})}{\alpha^2+1}\right)^2\right] \end{aligned} \tag{2.8}$$

[5]Mild conditions are irreduciblity and aperodicity which are satisfied by our transition kernels.

[6]I cannot use the block sampler suggested by Carter and Kohn (1994), which samples an entire series of latent volatility factor, since our model is nonlinear.

for $t = 2, \cdots, n-1$. Here and elsewhere in the paper, I use $X_{\backslash t}$ to denote $(x_j)_{j \neq t}$. [7]

Next, I move on the posterior distribution of parameters. Here I need their prior distributions. I assume the μ, ν and λ have Gamma distribution due to the positiveness and x_0 and α follows normal distribution and ρ has uniform distribution. Then the posterior distributions are given by,

(2.9)
$$p(x_0|X, Y, \beta, \alpha, \rho) \propto \frac{1}{\sqrt{2\pi \frac{1}{\alpha^2 + 1/\sigma_{x0}^2}}} \exp\left[-\frac{\alpha^2 + 1/\sigma_{x0}^2}{2} \left(x_0 - \frac{\alpha x_1 + \frac{\mu_{x0}}{\sigma_{x0}^2}}{\alpha^2 + 1/\sigma_{x0}^2} \right) \right]$$

(2.10)
$$p(\beta|X, Y, x_0, \alpha, \rho) \propto \left(\prod_{t=1}^{n} f(x_t, \beta)^{-\frac{1}{2}} \right) \exp\left[-\frac{\sum_{t=1}^{n-1} (u_t^2 - 2\rho u_t v_{t+1})}{2(1-\rho^2)} - \frac{1}{2} u_n^2 \right] p(\beta)$$

(2.11)
$$p(\alpha|X, Y, x_0, \beta, \rho) \propto \exp\left[-\frac{\sum_{t=1}^{n-1} (v_{t+1}^2 - 2\rho u_t v_{t+1})}{2(1-\rho^2)} - \frac{1}{2} v_1^2 \right] p(\alpha)$$

(2.12)
$$p(\rho|X, Y, x_0, \beta, \alpha) \propto (1-\rho^2)^{-\frac{n-1}{2}} \exp\left[-\frac{\sum_{t=1}^{n-1} (u_t^2 - 2\rho u_t v_{t+1} + v_{t+1}^2)}{2(1-\rho^2)} \right] p(\rho),$$

where $x_0 \sim \mathbb{N}(\mu_{x0}, \sigma_{x0}^2)$.

If I can draw the sample directly from (2.8) $\sim$ (2.12), then I complete all the algorithm for Gibbs Sampling. However, this is not the case except for (2.9) be-

---

[7] The conditional densities of the first and last latent factors are given respectively by

$$p(x_1|X_{\backslash n}, Y, \theta) \propto f(x_1, \beta)^{-\frac{1}{2}} \exp\left(-\frac{u_1^2 - 2\rho u_1 v_2}{2(1-\rho^2)} \right) \exp\left[-\frac{\alpha^2 + 1 - \rho^2}{2(1-\rho^2)} \left(x_1 - \frac{\alpha(x_2 + (1-\rho^2)x_0)}{\alpha^2 + 1 - \rho^2} \right)^2 \right]$$

$$p(x_n|X_{\backslash n}, Y, \theta) \propto f(x_n, \beta)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} u_n^2 \right) \exp\left[-\frac{(x_n - \rho u_{n-1} - \alpha x_{n-1})^2}{2(1-\rho^2)} \right]$$

cause of the complexity of the density function.[8] Then I may rely on Metropolis-Hastings (MH) sampling. The reader is referred to Chib and Greenberg (1995) for more detail MH sampling. In MH sampling from state-space model, the choice of proposal density function matters in terms of efficiency. Geweke and Tanizaki (2001) shows that when I sample latent factors a density function obtained from the transition equation is most efficient and so are priors when I sample parameters. So I choose prior density functions as proposal density function for parameters and $\frac{\sqrt{\alpha^2+1}}{\sqrt{2\pi(1-\rho^2)}}\exp\left[-\frac{\alpha^2+1}{2(1-\rho^2)}\left(x_t-\frac{\rho u_{t-1}+\alpha(x_{t-1}+x_{t+1})}{\alpha^2+1}\right)^2\right]$ as proposal density functions for latent factors.[9]

2. Simulation Study of Gibbs Sampling

Since the Gibbs sampling takes much longer time than ML estimation, I cannot do Monte-Carlo Experiments for Gibbs sampling here. But I do several treatment test, for example, different number of observations, different realization of error terms. They all give similar performance. So I report one example. True parameters and assumption on priors are shown in Table 11. I run 100000 iterations and discard first 36000 samples for burn-in period, which is determined by convergence diagnostics suggested by Geweke (1992). Convergence Diagnostic (CD) is given by

$$CD=\frac{\overline{\theta}_A-\overline{\theta}_B}{\sqrt{\frac{\hat{f}_A(0)}{n_A}+\frac{\hat{f}_B(0)}{n_B}}}$$

---

[8] I can draw x_0 from $\mathbb{N}\left(\frac{\alpha x_1+\frac{\mu_{x0}}{\sigma^2_{x0}}}{\alpha^2+1/\sigma^2_{x0}},\frac{1}{\alpha^2+1/\sigma^2_{x0}}\right)$.

[9] For proposal density function at t=1, $\frac{\sqrt{\alpha^2+1-\rho^2}}{\sqrt{2\pi(1-\rho^2)}}\exp\left[-\frac{\alpha^2+1-\rho^2}{2(1-\rho^2)}\left(x_1-\frac{\alpha(x_2+(1-\rho^2)x_0)}{\alpha^2+1-\rho^2}\right)^2\right]$ and for t=n, $\frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left(-\frac{(x_n-\rho u_{n-1}-\alpha x_{n-1})^2}{2(1-\rho^2)}\right)$.

Table 3.: Simulation Result with Gibbs Sampling.

Parameters	True Value	Priors	Posterior		Convergence
			Mean	Std. dev.	Diagnostics
μ	1	$G(2,1)$	1.1263	0.1793	0.9836
ν	3	$G(2,10)$	2.0719	0.3761	0.8039
λ	0.3	$G(2,\frac{1}{4})$	0.6773	0.3260	-0.3469
α	0.99	$N(0.9,0.05^2)$	0.9830	0.0063	-0.6327
ρ	-0.5	$U[-1,1]$	-0.3963	0.1212	-1.8713
x_0	-5	$N(-5,2^2)$	-5.2218	1.8699	-0.6374

Notes: $G(a,b)$ denotes gamma distribution with mean ab and variance ab^2 and $N(a,b^2)$ denotes normal distribution with mean a and variance b^2 and $U[a,b]$ denotes uniform distribution with support of $[a,b]$.

where A is the set of Gibbs samples with n_A iterations after burn-in period and B is the set of Gibbs samples with last n_B observations and f(0) is the spectral density at the zero frequency with Parzen window.[10] By the convention I set $n_A/n = 0.1$ and $n_B/n = 0.5$ where n denotes the number of Gibbs samples after burn-in periods.[11] If the sequence of Gibbs samples for a parameter is stationary, CD converges to the standard normal distribution as the number of samples goes to the infinity. Since the absolute value of CDs for all the parameters are less 1.96, they can be considered to pass the stationary test.

The results of the Gibbs sampling are summarized in Table 11. It shows that all the parameters except ν can be estimated significantly and unbiasedly, which is very similar implication for its slow convergence as Monte-Carlo experiments in Section C.2.

Figure 3 plots the series of extracted latent factor versus those of true latent

[10]In the Geweke (1992), the denominator does not have the square root term. Then it becomes average long-run variance.

[11]I need $n_A/n+n_B/n < 1$, if the ratios n_A/n and n_B/n are fixed, to get asymptotical independence between A and B as n goes to the infinity. See the Geweke (1992).

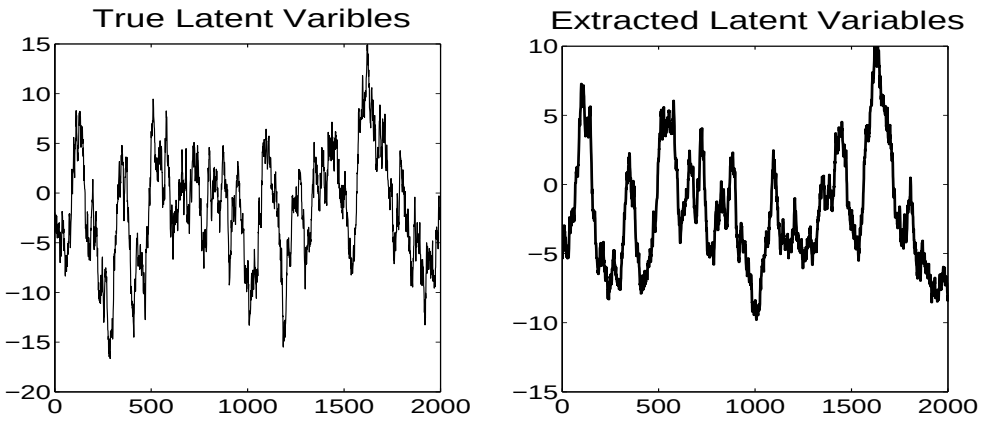

Fig. 3.: True and Extracted Latent Variables from the ML Estimation.

factor. It shows that the series of extracted variables follows true series closely.

E. Empirical Applications

1. Data Description

The series I use for application are the S&P 500 index weekly return series from the first week of December 1970 to the first week of January 2008 and monthly growth series of the dividend from February 1959 to December 2006. S&P 500 is obtained from website (http://finance.yahoo.com) and Dividend data from CRSP. S&P 500 weekly data is generated by Monday-to-Monday returns while dividend growth data is generated from seasonally-adjusted real dividend series which is created using value-weighted returns with dividend and without dividend.[12] S&P500 return is in natural numbers and dividend growth is in percentage. There are altogether 1,938 and 575 observations, respectively. Table 10 provides summary statistics for two series. In order to make the estimation more reliable and stable, I do three treatments to the

[12]Real dividend = (returns with dividend - returns without dividend) * market price index / CPI.

Table 4.: Summary Statistics.

	S&P 500		
	Original	After treatment	Dividend
Mean	0.0014	0.0000	0.0000
Max	0.1320	0.0832	8.2017
Min	-0.1301	-0.1185	-12.8766
Std. Dev.	0.0214	0.0210	3.2099
Skewness	-0.3724	-0.3919	-0.7182
Kurtosis	6.2128	5.1543	4.3462
Mean of $\lvert y_t \rvert$	0.0161	0.0159	2.3765

data. First, I subtract the non-zero mean because non-zero mean is not consistent with zero-mean error terms in the measurement equation. Secondly, I remove the data which are defined as more than 6 times of standard deviations upward or downward movement. These events may be considered as "outliers", which are independent of other observations in the sample.[13] Finally, I divide y_t by the average of absolute value of series. It is nothing but change in the assumption of unit variance of u_t. Hence, theoretically it changes the estimation results only up to scaling μ and ν. However, I find sometimes that whole estimation fails with raw data. I conjecture that so small numbers in the exponential function of normal density makes probability close to zero.

2. Stock Returns and Volatility Factor

Our plan is following: First, I do the ML estimation with the possibility of the stationary latent factor and correlation between errors. Secondly, if the latent factor is found to be nonstationary, I do the ML estimation setting $\alpha = 1$ or if the correlation of errors is found to be insignificant, I do the ML estimation setting $\rho = 0$ or both.

[13]For S&P 500 series, there are 2 observations removed, which are oil shock in 1974, black Monday in 1987. But for dividend growth series, there are no outliers.

Table 5.: ML Estimation Results with S&P 500 Returns.

	I(0) or I(1)		I(1)	
Parameters	**Leverage**	No leverage	Leverage	No leverage
ρ	**-0.5309***	0	-0.4773*	0
	(0.0768)		(0.0866)	
α	**0.9805***	0.9926*	1	1
	(0.0105)	(0.0066)		
μ	**0.00012***	0.00015*	0.00013	0.00013*
	(0.00003)	(0.00003)	(0.00010)	(0.00003)
ν	**0.00115***	0.00136*	9202.2	40.4496
	(0.00030)	(0.00088)	(2.09e7)	(11325.9)
λ	**0.4100***	0.3315*	0.2152	0.1869*
	(0.1133)	(0.1422)	(0.1698)	(0.0426)
x_0	**-11.2393***	-16.1677	-87.8777	-74.3899
	(5.3607)	(7.4214)	(1.05e4)	(1.50e3)

Notes: parenthesis denotes standard errors and asterisk mark denotes significance at 5 percent level. Bold column stands for the best estimation result.

However, I report ML estimation result with all four cases regardless of the result of the first trial if possible. Secondly I do the Gibbs sampling with best assumption sets chosen from ML estimation and compare it with ML estimation results. Though I try two set of assumptions which are either setting $x_0 = 0$ or setting $\kappa = 0$, I report the result with setting $\kappa = 0$ because I think latter more reliable. This plan is applied to next section also.

The Table 5 shows the ML estimation results of Dow Jones stock return with various assumptions. All the results are estimated with setting $\kappa = 0$ because I cannot get significant results from setting $x_0 = 0$.[14] The second column tells that latent factor

[14]Under the stationary latent factor case setting $x_0 = 0$ and setting $\kappa = 0$ are not neutral selection because the effect of the former dies out eventually but the effect of the latter does not. However, setting $\kappa = 0$ is not weird in the sense that the mean of the latent factor becomes the same as the reflection point and hence, I may see equal possibility of high or low volatility regime.

Table 6.: Gibbs Sampling Results with S&P 500 Returns.

Parameters	Priors	Posterior		Convergence
		Mean	Std. dev.	Diagnostics
ρ	$U[-1,1]$	-0.4173	0.0874	-1.8114
α	$N(0.9, 0.05^2)$	0.9840	0.0063	-1.7800
μ	$G(2,1)$	0.00014	0.00003	-0.6189
ν	$G(2,10)$	0.00112	0.00021	-0.3290
λ	$G(2,\frac{1}{4})$	0.4358	0.1143	-0.8595
x_0	$N(-11,2^2)$	-6.3075	1.7578	-1.1912

Notes: $U[a,b]$ denotes uniform distribution with a support (a,b) and $G(a,b)$ denotes gamma distribution with mean ab and variance ab^2 and $N(a,b)$ denotes normal distribution with mean a and variance b.

is near unit-root process though it does not have exact unit root.[15] And I confirm the leverage effect because the correlation coefficient is negative and significant as I expected. The quantification of this leverage effect is already presented at Table 2 in the Section B. The minimum and maximum level of volatility can be inferred from $\hat{\nu}$ and $\hat{\nu} + \hat{\mu}$ multiplied by mean of $|y_t|$ as shown in the last column which is obtained from the second column if I set $\sigma = 1$. So the low volatility level is $\sqrt{0.0001} * 100 = 1.1$ percent while the high volatility level is $\sqrt{0.0011 + 0.0001} * 100 = 3.6$ percent in a week. The slope parameter, $\hat{\lambda}$ indicates relative slow transition because it takes at least 13 weeks to move from low volatility regime, say $x_t = -3$ to high volatility regime, say $x_t = 3$ even with successive positive shock of $v_t = 0.5$. From the third to fifth columns indicates that more restrictions make the estimation worse. So hereafter I use the second column.

Secondly, I run Gibbs sampling with MH algorithm. I samples 90000 iterations and discard 32000 iterations based on convergence diagnostics as before. CD is reported in the Table 6. Figure 4 shows Gibbs samples after burn-in period. CD

[15] t-statistic of ADF unit root test for extracted latent factor is -4.90 while the critical value of 5 % confidence level is -2.86.

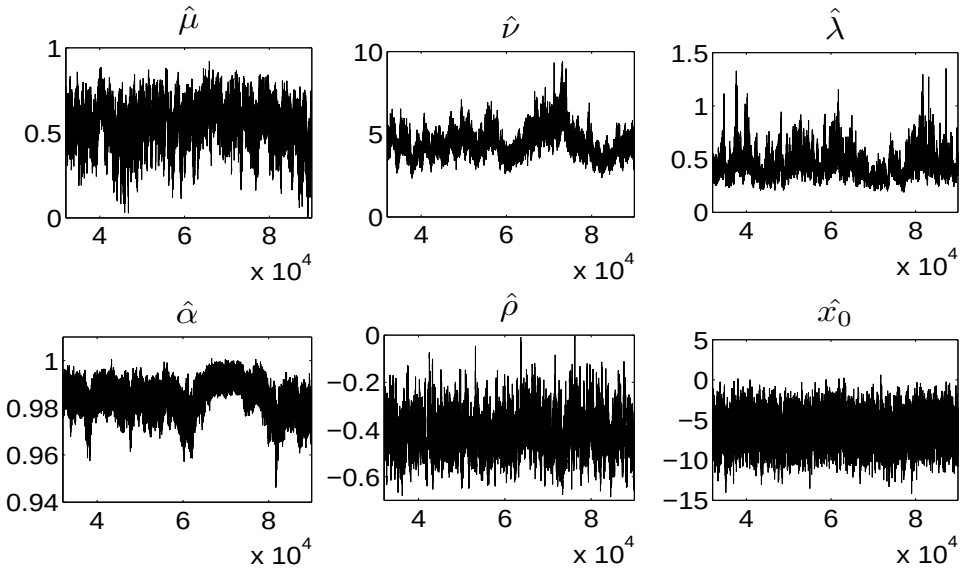

Fig. 4.: Gibbs Samples with S&P 500 Return.

statistics and Figure 4 indicates the Gibbs sampling converges well and hence is reliable. Table 6 shows that the posterior mean is very close to the ML estimators. Fifth column is the inferred posterior mean with $\sigma = 1$.

I extract the latent factors from two alternative method and find that they are very similar and seems that they share the common stochastic component as illustrated in Figure 5. Straight line in the Figure 5 can be thought as a threshold for the regime change in a rough sense. The area above the upper straight line is high volatility period and vice versa. So I can interpret that the economy usually stays in the low volatility regime or transition period while it have stayed in the high volatility regime in mid-70s and around 2000 for some time and it almost entered 80s couple of times. And recently the economy moves towards high volatility regime very rapidly. Hereafter I are going to use the series of the latent factors or volatility factor from ML estimation though the all the remaining results are very similar with Gibbs sampling results.

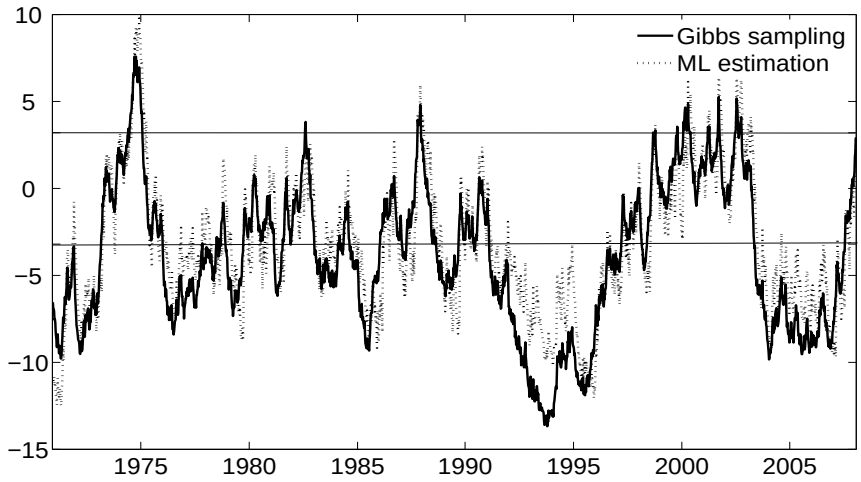

Fig. 5.: Volatility Factors with S&P 500 Returns.

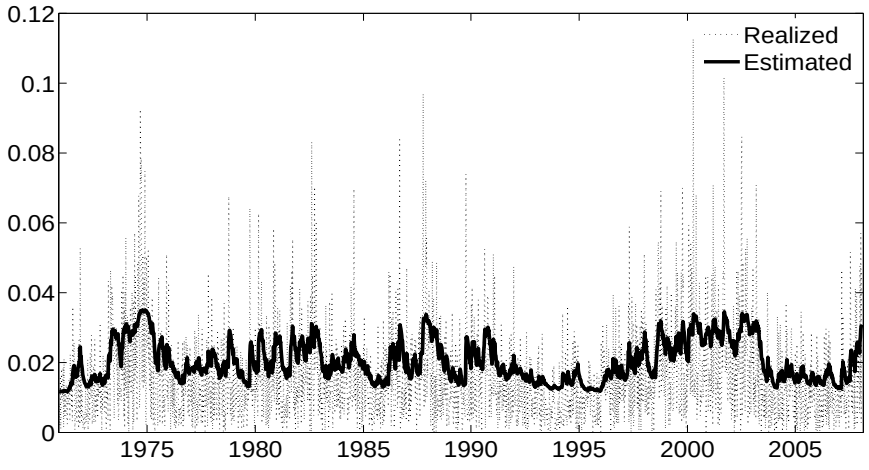

Fig. 6.: Comparison between the Realized and the Estimated Volatility.

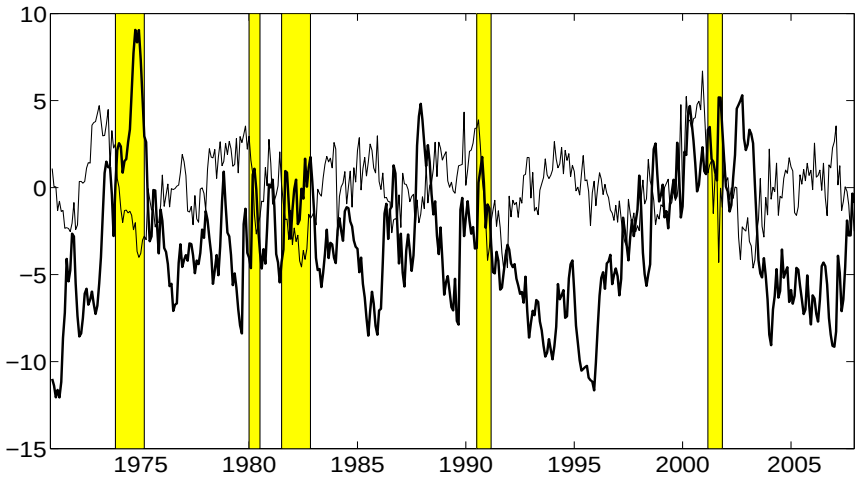

Fig. 7.: Business Cycle and Volatility Factor.

Figure 6 plots the estimated volatility component with parameter obtained from ML estimation and latent factors. The discrepancy between realized volatility, $|y_t|$ and estimated volatility $E[\sqrt{f(x_t)}|\mathcal{F}_t]$ can be considered as the realization of error term of N(0,1), u_t. Generally, however, estimated volatility explains well the $|y_t|$. We also see that estimated volatility has a tendency to resist to big shocks quite well. If large volatility is a realization of an outlier and the outlier effect will disappear at the next period, the smooth transition regimes models in contrast to the conventional stochastic volatility models would not be affected much so that this resistance will increase the accuracy of forecasting volatility of the next period.

Finally, I try to link the volatility factor with some macro-economic fundamentals. Figure 7 shows the relationship between monthly real expenditure on nondurable goods and services and the volatility factor. The data is obtained from St. Louis Fed and hp-filtered. Monthly volatility factor are obtained from simple average of weekly data. I see that for 40 years reversed volatility factor follows the GDP series very

closely except 1998 $\sim$ 2001.[16] Moreover, as So, Lam, and Li (1998) and Hamilton and Lin (1996), economic recessions defined by NBER are clearly associated with high volatility periods. For 40 years, when the US economy suffered 5 recessions volatility factor shows trough. And though late 80s is not defined as recession period, GDP series sharply decreases with volatility factor.

While the volatility factor is not unit-root process, I can focus on only after 90s to make it nonstationary. I do the Variable Addition Test of cointegration for two series suggested by Park (1990). I add superfluous four trend in the cointegrating regression and test the significance of the coefficients of these irrelevant regressors because they are all zeros under the null hypothesis, which is two series are already cointegrated. I construct the cointegrating regression with structural break to take care of reverse relation during 1998 $\sim$ 2001. The test indicates two series are cointegrated.[17]

3. Dividend Growth and Volatility Factor

The Table 7 shows ML estimation results with several cases of the assumptions. The second column tells that the latent factor follows clearly unit-root process. Actually the unit root test for the extracted latent factor shows the nonstationarity.[18] Moreover, I see that the correlation between the measurement error and the transition error are big positive and significant. But this means that when the dividend growth at current period increases the volatility of the dividend growth at the next period is more likely to increase. The plausible explanation of this reverse leverage effect is

[16]At that time the economy is in good shape though the stock market becomes volatile due to Asian currency crisis and burst of IT bubble.

[17]Test statistics are 3.31. Under the null hypothesis test statistics follows $\chi^2(4)$. Since the p-value of 3.31 is 0.5067 I cannot reject the null hypothesis.

[18]t-statistic of ADF unit root test is -2.49 while the critical value for 5 % confidence level is -2.87.

Table 7.: Estimation Results with Dividend Growth.

	I(0) or I(1)		**I(1)**	
Parameters	Leverage	No leverage	**Leverage**	No leverage
ρ	0.7228*	0	**0.7228***	0
	(0.0794)		(0.0794)	
α	1.0000	0.9948*	**1**	1
	(n.a.)	(0.0057)		
μ	1.0266*	1.1911*	**1.0266***	1.2329*
	(0.5367)	(0.7037)	(0.5367)	(0.6212)
ν	57.1422	40.7222*	**57.1429**	45.0354*
	(34.8674)	(15.2283)	(34.8691)	(19.8046)
λ	0.2751*	0.2928*	**0.2751***	0.2697*
	(0.0948)	(0.1097)	(0.0948)	(0.0997)
x_0	-18.5818	-15.1219	**-18.5818***	-15.6723*
	(6.7331)	(5.6189)	(6.7333)	(6.1586)

Notes: parenthesis denotes standard errors and asterisk mark denotes significance at 5 percent level. Bold column stands for the best estimation result.

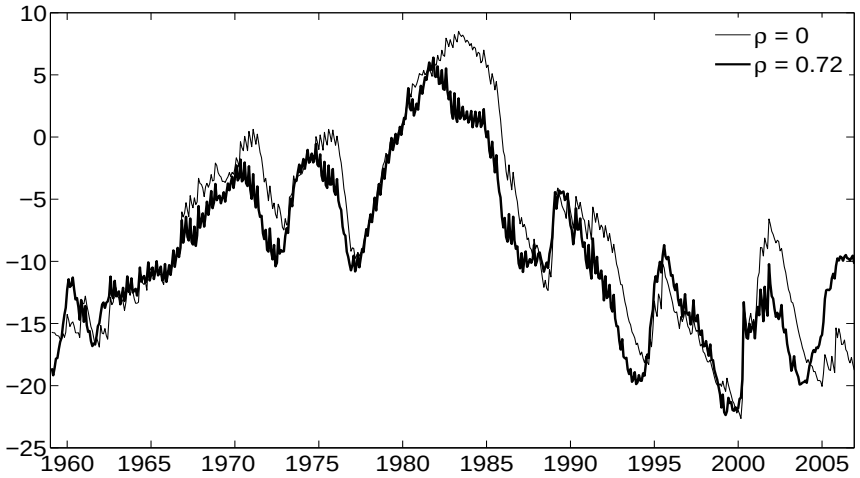

Fig. 8.: Volatility Factors with and without Restriction.

Table 8.: Conventional Model Estimation Results with Dividend Growth.

	I(0) or I(1)		**I(1)**	
Parameters	Leverage	No leverage	Leverage	No leverage
ρ	0.7038*	0	0.6982*	0
	(0.0801)		(0.0749)	
α	0.9935*	0.9891*	1	1
	(0.0054)	(0.0073)		
ν	10.7966	6.2689	1.2216	1.9885
	(9.6089)	(3.7873)	(n.a.)	(234.053)
λ	0.1621*	0.1535*	0.1593*	0.1403*
	(0.0283)	(0.0302)	(0.0268)	(0.0271)
x_0	-14.6614*	-9.0032	-0.2370	-0.3793
	(7.0245)	(5.8110)	(n.a.)	(839.08)

Notes: Parenthesis denotes standard errors and asterisk mark denotes significance at 5 percent level.

beyond the purpose of this paper. Anyway our next candidate assumption is setting $\alpha = 1$ as shown in the fourth column. It tells that I have the same estimation result with nonstationary assumption. The result with restrictions of $\rho = 0$ is presented at third and fifth column for the comparison. I think the fourth column as the most promising result though it has one slightly insignificant estimate because I cannot ignore big correlation coefficient. But in terms of pattern of the latent factor series, the stochastic volatility models with smooth transition regimes are not much sensitive to this restriction as illustrated in the Figure 8. And the low volatility level is $\sqrt{1.0266} = 1.01$ percent while the high volatility level is $\sqrt{54.1429 + 1.0266} = 7.42$ percent in a month. The slope parameter, $\hat{\lambda}$ indicates very slow transition because it takes at least 20 months to move from low volatility regime, say $x_t = -5$ to high volatility regime, say $x_t = 5$ even with successive positive shock of $v_t = 0.5$.

I also try to estimate the conventional stochastic volatility model with dividend growth data as shown in Table 8. Since I have unit-root process of latent factors

Table 9.: Gibbs Sampling Results with Dividend Growth.

Parameters	Priors	Posterior		Convergence
		Mean	Std. dev.	Diagnostics
ρ	U[-1,1]	0.6570	0.0873	0.1205
α		1		
μ	$G(2, 0.1)$	0.9906	0.3903	2.2556
ν	$G(2, 10)$	42.1867	9.1226	1.5684
λ	$G(2, \frac{1}{4})$	0.3573	0.0645	1.8240
x_0	$N(0, 15^2)$	-12.0517	1.7420	-3.0078

Notes: $U[a, b]$ denotes uniform distribution with a support (a, b) and $G(a, b)$ denotes gamma distribution with mean ab and variance ab^2 and $N(a, b^2)$ denotes normal distribution with mean a and variance b^2.

without restriction[19], I impose nonstationarity. Then I do not have good estimation results because hessian is not positive definite so that some of estimates cannot have standard errors. I think that this is one example of the weakness of nonstationary conventional volatility models which have explosive volatility eventually as the number of observations increases.

Secondly, I run Gibbs sampling with MH algorithm. I samples 60000 iterations and discard 24000 iterations based on convergence diagnostics which is reported in the Table 9. Though the convergence diagnostics of posterior means of μ and x_0 are bigger than 2, Figure 9 supports the convergence of the Gibbs samples after burn-in period. Table 9 shows that the posterior mean is comparable to the ML estimators. In addition, each posterior mean is significant. So hereafter I use Gibbs sampling result as our final estimation.

I extract the latent factors from two alternative method and find that they are very similar and share the common stochastic component as illustrated in Figure 10. Straight line can be thought as a threshold for the regime change in a rough sense.

[19]The t-statistic of ADF unit root test for the extracted latent factor is -2.34 while the critical value for 5 % confidence level is -2.87.

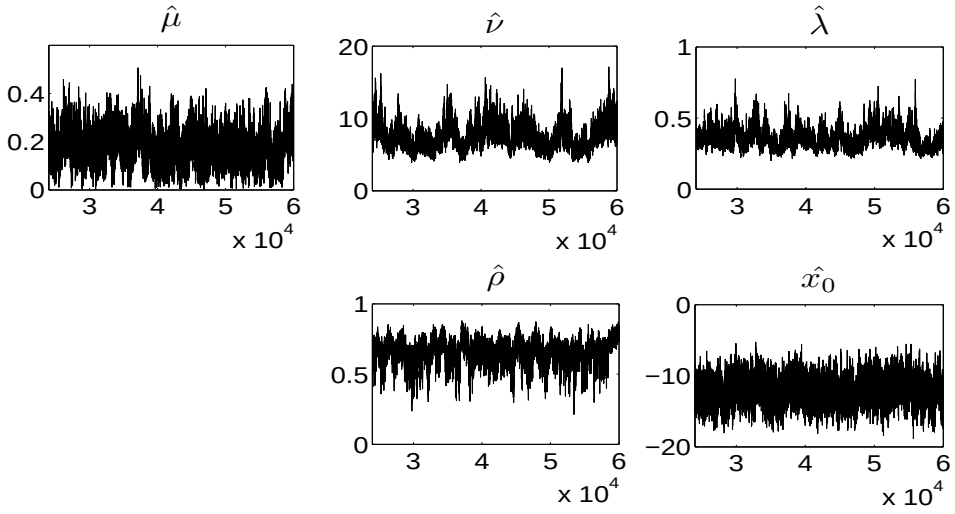

Fig. 9.: Gibbs Samples with Dividend Growth.

The area above the upper straight line is high volatility period and vice versa. So I can interpret that the fundamentals which generates volatility have stayed in the low volatility regime in 60s and moved into the transition period in 70s and entered into high volatility regime in first-half 80s. From late 80s until now the fundamentals have stayed in the low volatility regime.

Figure 11 plots the estimated volatility component with parameter obtained from ML estimation and latent factors. Estimated volatility from Gibbs samples is obtained by Monte-Carlo integration i.e.,

$$\begin{aligned} E[\sqrt{f(x_t)}|\mathcal{F}_n] &= \int \sqrt{f(x_t)} p(x_t|\mathcal{F}_n)\, dx_t \\ &\approx \frac{1}{G} \sum_{i=k+1}^{G} \sqrt{f(x_t^{(i)})} \end{aligned}$$

where $x_t^{(i)}$ is the i-th sample for x_t and k is the length of burnin period and G is the total number of Gibbs samples for x_t. I see that generally estimated volatility explains well realized volatility, $|y_t|$.

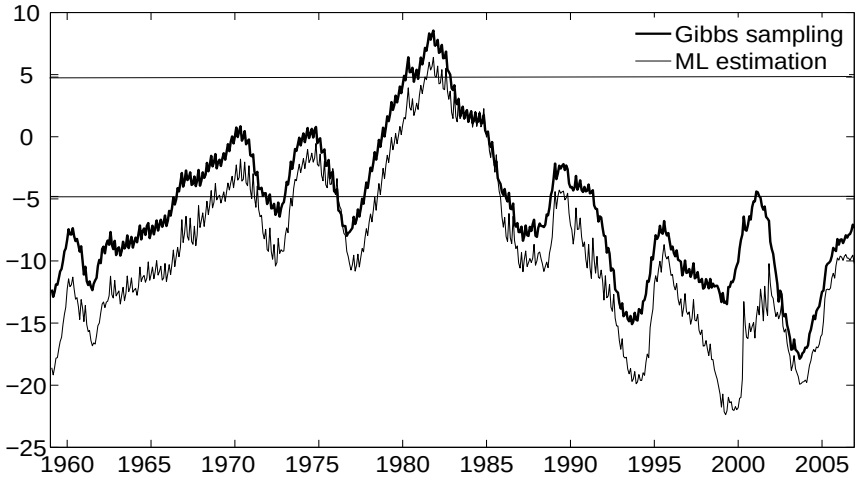

Fig. 10.: Volatility Factor with Dividend Growth.

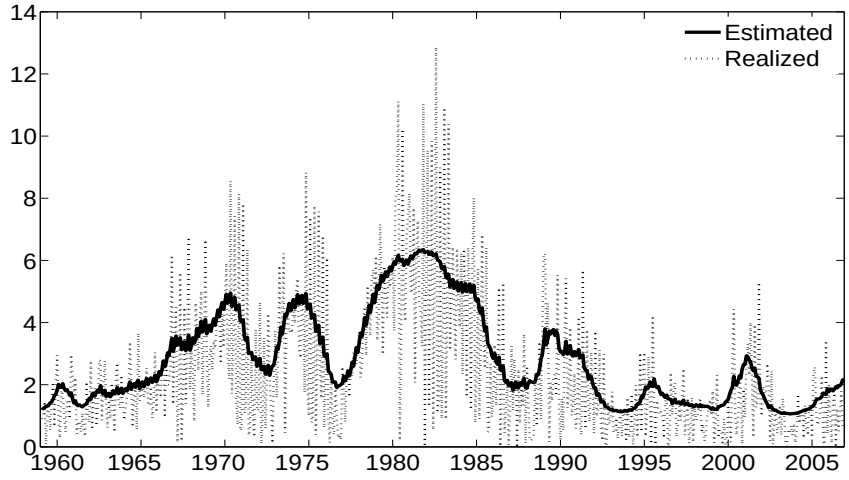

Fig. 11.: Comparison between the Realized and the Estimated Volatility.

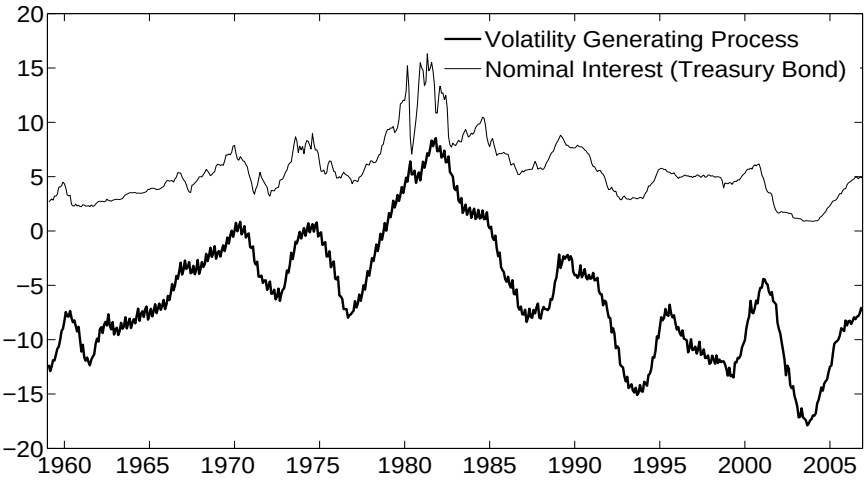

Fig. 12.: Volatility Factor and Nominal Interest Rates.

Finally, I try to match the volatility factor with some macro-economic fundamentals. Figure 12 shows the relationship between monthly nominal interest rates and the volatility factor. The interest rates data is obtained from 3-month treasury bond at St. Louis Fed. Monthly nominal interest rates and the volatility factor seems to share the common stochastic trend and the cointegration test confirms it.[20]

F. Conclusion

I construct nonlinear state-space model with persistent latent factors to investigate the volatility factor from the return or growth data of an asset. I argue that the stochastic volatility models with smooth transition regimes have some advantages over the conventional models for the implication to the non-explosive volatility and asymmetric leverage effect. Due to the nonlinearity of logistic volatility function

[20]The t statistic of the residual-based cointegration test is -3.31 which means the power of reject null hypothesis is slightly weak. But the Johansen Cointegration Test confirms the cointegration relationship. p-value with Maximum Eigenvalue is 0.0290 and with Trace 0.0104.

and persistence of the latent factors conventional Extended Kalman filtering is not appropriate with our model. To filter the volatility factor from data I introduce two alternative methods, which are density-based ML estimation and Gibbs sampling. These two methods give similar results though only density-based ML estimation allows extensive Monte-Carlo Experiments. I find that all the parameter can be correctly estimated from Monte-Carlo Experiments. I apply our methods to stock return and dividend growth data and extract the volatility factor. The extracted volatility factor explain the realized volatility quite well and moreover I can find some fundamentals cointegrated with them. One can extend our model into multivariate setup for extracting common stochastic trend from several data sets. Though I can use both methods in solving the multivariate model again, I expect that density-based ML estimation will have a limitation due to the dimension.

CHAPTER III

MACROECONOMIC UNCERTAINTY AND ASSET PRICES: A STOCHASTIC VOLATILITY MODEL

A. Introduction

Time-varying macroeconomic uncertainty is an important ingredient for asset valuation. Due to the nature of aggregate shocks, macroeconomic uncertainty is reflected in equilibrium asset prices because asset holders will demand some premium for bearing such undiversified risk. However, macroeconomic uncertainty is not an observable to economists, and therefore modelling and measuring this uncertainty is a meaningful, yet challenging task. Furthermore, the most popular macroeconomic asset pricing models identify consumption growth process as the link between macroeconomic variables and asset returns, exploiting the simple and elegant Euler equation of consumption growth and asset returns. Alas, aggregate consumption is close to a random walk and the size of unconditional variance of consumption growth is fairly modest to justify high average equity premium with low and stable interest rates. These puzzles based on consumption based asset pricing models have been one of the main research questions in finance and macroeconomics since Hansen and Singleton (1982) and Mehra and Prescott (1985). In this paper, I tackle this issue by measuring time-varying uncertainty of macroeconomic variables and studying the links to market returns via estimating a consumption-based asset pricing model with a non-expected utility function. As the first step, I develop a stochastic volatility model and propose an econometric procedure to extract common and idiosyncratic volatility for multivariate processes. Novel features of our volatility model include a unit root common factor and logistic volatility function. This setup allows a persistent conditional

volatility shifting between two regimes (high and low uncertainty regimes) with a smooth transition. The existence of transition period between high and low volatility regimes implies that economic agents may dislike transition periods because of uncertainty in regimes that they belong to. Epstein and Zin (1989) show that economic agents prefer an earlier resolution of uncertainty if risk aversion parameter is bigger than the reciprocal of elasticity of intertemporal substitution in case of Kreps-Porteus utility function. Thus, together with this non-expected utility function, our volatility setup can generate a higher risk premium even with medium level of volatility.

In terms of econometric setup, our volatility model can be regarded as a non-linear, non-stationary state space model. There are several non-linear filtering techniques in the conventional approach to solve such a stochastic volatility model with nonlinear measurement equation. But multi-dimensionality of our problem makes these filtering techniques much more difficult to be applied. To overcome this issue, I use Gibbs sampling which does not suffer seriously from the curse of dimensionality, because it utilizes univariate conditional density function. Here I develop an algorithm to filter macroeconomic uncertainty based on the chapter II which studies the univariate stochastic volatility model with a logistic function.

When I put our methods into data, since I am interested in performances of consumption based asset pricing model, I extract a common and idiosyncratic volatility factors for consumption and dividend growth. I find that the common factor delineating 'macroeconomic uncertainty' predicts post-war business cycle recessions quite well. Then, I estimate a long-run risk model of asset prices incorporating this macroeconomic uncertainty. According to our estimation, both risk aversion and the intertemporal elasticity of substitution are estimated around two, and our simulation results show that the model can match the first and second moments of market return and risk-free rate, hence explains the equity premium, the risk-free rate puzzle, and

volatility puzzle.

Our paper is related to at least two strands of literatures. First and most direct line of literature is the long-run risks asset price models. Bansal and Yaron (2004) set consumption and dividend growth processes to contain a small, but persistent process as well as a stochastic volatility term and show that they can explain many stylized facts in asset market. In their paper, they emphasize the long-run risks channel, which is based on the common portion of conditional expectations that move slowly over time. Mainly due to its persistence, this model can generate sufficiently high risk premium because equity price is a discounted sum of all the future dividends. However, the role of stochastic volatility in their model is not directly related to accounting for equity premium puzzle, risk-free rate puzzle, and volatility puzzle, but mostly for explaining time-varying risk premium. I extend this model to include a more realistic volatility setup to analyze how important this volatility channel is. According to their recent empirical paper (Bansal, Kiku, and Yaron (2007), the estimates of risk aversion and the elasticity of intertemporal substitution are around 10 to 15, and 0.5 respectively with or without a stochastic volatility setup. Our results suggest that a more flexible and realistic stochastic volatility can substantially improve the performance of long-run risks model, while generating sufficient size of equity premium and its volatility, as well as low and stable interest rate. Another important asset pricing model is the habit formation preference. For instance, Campbell and Cochrane (1999) can generate equity premium with low and stable interest rate via time-varying risk aversion across the state of the economy. A drawback of this approach is the average risk aversion should be very high to explain equity premium as Campbell (2002) pointed out. Also, there exist many papers by relaxing other environment setting while holding simple utility function assumption to address asset market behavior. For example, the existence of heterogenous agents in the sense that some people can-

not participate in the asset market increase the equity premium (See Constantinides, Donalson, and Mehra (2002)) or transaction costs makes return on risk-free bond low because it can generate utility like cash or rare disaster or event possibilities makes equity riskier.(See Barro (2005), Rietz (1988). See Campbell (2002) for the compact review).

The rest of the paper is organized as follows. In Section B, I develop our asset pricing model. Then, I introduce the methodology to identify macro uncertainty in section C. Section D shows the empirical results on equity premium. Then I conclude.

B. Asset Pricing Model

I consider a simple closed economy in which a representative agent has a Epstein-Zin-Weil recursive preferences (Epstein and Zin (1989) and Weil (1989)) given by

$$U_t = \left[(1-\delta)C_t^{\frac{1-\gamma}{\chi}} + \delta(E_t U_{t+1}^{1-\gamma})^{\frac{1}{\chi}}\right]^{\frac{\chi}{1-\gamma}}$$

where $\chi = \frac{1-\gamma}{1-\frac{1}{\psi}}$, $\gamma \geq 0$ is the coefficient of relative risk aversion, $\psi \geq 0$ is the intertemporal elasticity of substitution (IES) and $0 < \delta < 1$ is the time discount factor. Compared with a power utility function, Epstein-Zin-Weil utility function allows more flexibility because it breaks the tight link between the parameters risk aversion γ and intertemporal substitution ψ. For instance, Epstein-Zin-Weil preference can have both parameters are bigger than one, implying $\chi < 0$ holds. In case of power utility function, IES is a reciprocal of risk aversion parameter. Another useful property of this preference function is about resolution of uncertainty. If $\gamma > \psi^{-1}$, then economic agents prefer earlier resolution of uncertainty. The intertemporal budget constraint for the representative agent can be written as $W_{t+1} = R_{w,t+1}(W_t - C_t)$ where W_t is the wealth and $R_{w,t+1}$ is the gross return on the portfolio of all invested wealth on

consumption claims. Epstein and Zin use dynamic programming to derive an Euler equation for gross return $R_{i,t+1}$ for any asset i

$$1 = E_t \left[\delta^{\chi} \left(\frac{C_{t+1}}{C_t} \right)^{-\frac{\chi}{\psi}} R_{w,t+1}^{\chi-1} R_{i,t+1} \right] \tag{3.1}$$

From Euler equation (3.1) I have the logarithm of the intertemporal marginal rate of substitution (IMRS)

$$m_{t+1} = \chi \log \delta - \frac{\chi}{\psi} g_{c,t+1} + (\chi - 1) r_{w,t+1} \tag{3.2}$$

where $r_{w,t+1} = \log R_{w,t+1}$ is the log return on the portfolio of all invested wealth. Using the definition of $R_{w,t+1} = \frac{C_{t+1}+P_{w,t+1}}{P_{w,t}}$ and applying the log linearization method as in Campbell and Shiller (1988), I can find an approximate relationship between unobservable $(r_{w,t+1})$ and observable $(z_{c,t+1})$ and $(g_{c,t+1})$

$$r_{w,t+1} \approx k_0(\bar{z}_c) + k_1(\bar{z}_c) z_{c,t+1} - z_{c,t} + g_{c,t+1} \tag{3.3}$$

where $z_{c,t} = \log(P_{c,t}/C_t)$ is the log price-consumption ratio.[1]

To solve the model further I need to specify the process of $(z_{c,t})$. I assume the consumption growth, $g_{c,t+1}$ and dividend growth, $g_{d,t+1}$ have a long run component and a time-varying volatility component driven by scalar processes $x_{c,t}$ and $x_{d,t}$ respectively. Additionally, I assume that both of the volatility generating processes share a scalar common volatility generating process, or I call the macroeconomic uncertainty process, w_t. That is, I extend Bansal and Yaron (2004) by making more

[1]Coefficients of this approximation are as follows: $k_0(\bar{z}) = log(1+\exp(\bar{z})) - k_1(\bar{z})\bar{z}$, $k_1(\bar{z}) = \frac{\exp(\bar{z})}{1+\exp(\bar{z})}$ for some value $\bar{z}$.

realistic assumptions on volatility setup. Specifically, I have

$$g_{j,t+1} = \mu_j + \eta_{j,t} + \sqrt{f_j(x_{j,t})}\,\epsilon_{j,t+1} \tag{3.4}$$

$$\eta_{j,t+1} = \rho_j \eta_{j,t} + \varphi_j \sqrt{f_j(x_{j,t})}\,e_{j,t+1} \tag{3.5}$$

$$x_{j,t} = \lambda_j\, w_t + \nu_{j,t} \tag{3.6}$$

$$w_t = w_{t-1} + u_t \qquad \text{for} \quad j = c, d \tag{3.7}$$

$$f_j(x_{j,t}) = \alpha_j + \frac{\beta_j}{1 + \exp\left[-(x_{j,t} - \kappa_j)\right]} \tag{3.8}$$

where $\alpha_j > 0, \beta_j > 0$.

I assume that the error terms are characterized by

$$\begin{pmatrix} \epsilon_{c,t} \\ \epsilon_{d,t} \end{pmatrix} \sim iid\ \mathbb{N}\left(0, \Sigma\right), \quad \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{3.9}$$

and $(\nu_{j,t})$ is $i.i.d.$ $\mathbb{N}(0, \sigma_j^2)$, (u_t) and $(e_{j,t})$ are $i.i.d.\mathbb{N}(0,1)$. I assume that $(\varepsilon_{j,t}), (e_t)$, $(u_{t,})$ and $(\nu_{j,t})$ are mutually independent of each other except (3.9). The unit variance of $\epsilon_{j,t}$, $e_{j,t}$, and u_t is identifying restriction because α_j, β_j, φ_j and λ_j captures the variance of error terms. Furthermore, w_0 is assumed to be independent of (u_t), $(\nu_{j,t})$ and $(\epsilon_{j,t})$. $\mu_c + \eta_{c,t}$ represents the conditional expectation of consumption growth and $\eta_{c,t}$ is the term capturing long-run risks in consumption. Bansal and Yaron (2004) and Hansen, Heaton, and Li (2005) showed that persistence of this process measured by ρ_c plays an essential role of generating equity premium with high unconditional volatility without increasing the first and second moments of the risk-free rate. This channel of 'long run risk' is still a non-trivial part of our asset pricing model, but I focus more on how asset returns are connected to aggregate uncertainty that varies over time. I now explain the setup in the below.

Our conditional volatility setup has several attractive features. First, volatility of a macroeconomic variable is assumed to be the sum of a common factor(w_t) multiplied by its factor loading (λ), and an idiosyncratic factor (v_t). Given the unobservable nature of conditional volatility of macroeconomic variables, this strategy usually enables researchers to identify factors of interest more sensibly, by exploiting common variations yet allowing individual deviations. Second, I impose that these latent factors vary according to a logistic function (3.8), and further assume that the common factor w_t follows a random walk process. Intuitively, this implies that our volatility can have two bounded regimes with a smooth transition and volatility is clustering due to its persistence.[2] Park (2002) shows that a model with asymptotically homogeneous functions of an integrated process has several nice statistical properties. First, the sample autocorrelations of the squared processes have the same random limit for all lags i.e., strong persistence. Secondly, the sample kurtosis has supports truncated on the left by the kurtosis of the innovations i.e., leptokurtosis. Since the logistic function belongs to the class of asymptotically homogeneous function, our model can capture the volatility clustering and fat-tail features of time-series data. Lastly, the smooth transition feature of our volatility model together with random walk common factor provides us with some insight about macroeconomic uncertainty. There is little doubt that high volatility regime is more uncertain than low volatility regime. However, if volatility is in between two regimes, economic agents may dislike this type of

[2]Note that existing models such as GARCH models cannot preclude explosive dynamics which is not so realistic given historic evidences. The parameters α_j and $\alpha_j + \beta_j$ represent two asymptotic levels, i.e., low volatility regime and high volatility regime, respectively. The assumption of positivity of β_j makes the logistic function upward sloping and hence, larger latent variable means higher volatility. This is not a restriction but for convenience, because the latent variable will be extracted reversely when a logistic function is downward sloping. The parameters κ_j and λ_j in the (3.6) specify the transition between two regimes, i.e., the reflection point and the speed of the transition, respectively.

Knightian uncertainty, because they are not sure about which regime they will end up with in the future. Thus, in comparison with conventional regime shifting models allowing only abrupt changes, our setup predicts that an economy with medium level of macroeconomic volatility (i.e. in a transition mode) might request a higher premium than otherwise, because of this additional uncertainty different than risk. To further analyze asset pricing implications of our model, I solve the log price-consumption ratio $z_{c,t+1}$ as

$$z_{c,t+1} \approx A_{0,c} + A_{1,c}\eta_{c,t} + A_{2,c}f_c(\lambda_c w_t). \tag{3.10}$$

The relevant state variables are $\eta_{c,t}$ and w_t. The solutions for $A_{0,c}$, $A_{1,c}$, and $A_{2,c}$ are in the Appendix A. If $\gamma > 1$ and $\psi > 1$, then $A_{1,c} > 0$ and $A_{2,c} < 0$. This means that consumers like higher expected future growth, but does not prefer a rise in macroeconomic volatility. In addition, I can see that when λ_c gets smaller, i.e. there exists a slower speed of adjustment, the absolute value of $A_{2,c}$ can increase. That is, given the traditional risk-return tradeoff relationship, the extent to which consumers dislike volatility may become larger because of slower resolution of uncertainty.

Once all the coefficients are verified, I can derive easily the innovations in IMRS

$$m_{t+1} - E_t(m_{t+1}) \approx \Lambda_{m\epsilon}\sqrt{f_c(x_{c,t})}\epsilon_{c,t+1} - \Lambda_{m,e}\sqrt{f_c(x_{c,t})}e_{c,t+1} - \Lambda_{m,u}(t)u_{t+1} \tag{3.11}$$

where the coefficient terms are defined in Appendix A. Here the risk sources are three terms, labeled as short run risk ϵ_{t+1}, long run risk e_{t+1} and common macroeconomic uncertainty u_{t+1}.

Next, I derive the market return. Although I solve most of our cases numerically thus I do not resort to this approximate solution when estimating the model, I write

down this formula with the derivation in the Appendix B for illustration purposes.

$$E_t[r_{m,t+1} - r_{f,t}] = \Lambda_{m,e} k_{1,m} A_{1,m} \varphi_c f_{c,t} + \Lambda_{m,u}(t) k_{1,m} A_{2,m} \left(\Gamma_1 f_{c,t} + \Gamma_2\right) - \frac{1}{2} Var_t(r_{m,t+1})$$

The first term in the right hand side is basically the contribution by long-run risks. The main focus of our paper is the second term, how macroeconomic uncertainty affects excess expected return for holding equity. It is noteworthy to see that the market price of the macroeconomic uncertainty is time-varying. In fact, the time-variability term in $\Lambda_{m,u}(t)$ coincides with the uncertainty factor. Therefore, when volatility is large, the effect on expected excess return can be bigger. Finally, I express the risk-free rate as follows:

$$\begin{aligned} r_{f,t} &= -\log\left[E_t \exp(m_{t+1})\right] \\ &= -\log\delta + \frac{1}{\psi} E_t g_{c,t+1} + \frac{1-\chi}{\chi} E_t\left[r_{w,t+1} - r_{f,t}\right] - \frac{1}{2\chi} Var_t(m_{t+1}) \end{aligned} \tag{3.12}$$

I also derive it in terms of current value state variables.(See the further derivation in Appendix C). With this formula (3.12), however I can see the comparative statics more easily. The negative relation between risk-free rate and IES is clear from (3.12). But as the risk-aversion coefficient becomes larger and hence, both the conditional mean of risk premium on wealth and the conditional variance of IMRS increase, the effect on the risk-free rate is not clear from (3.12) immediately because last two terms have opposite direction. If the third term is more dominant than the fourth term, the risk-free rates decreases and this fact is confirmed from the simulation.

C. Identifying Macroeconomic Uncertainty

1. A Bayesian Algorithm

In this section, I propose an econometric procedure to identify macroeconomic uncertainty process based on the model I developed in the previous section. For this purpose, I rewrite the measurement equation (3.4) as

$$y_{j,t+1} = \sqrt{f_j(x_{j,t})}\,\epsilon_{j,t+1} \tag{3.13}$$

where $y_{j,t+1} = g_{j,t+1} - \mu_j - \eta_{j,t}$.

Note that $y_{j,t}$ is net of unconditional and conditional mean terms. Thus, I need to subtract these component to identify $y_{j,t}$ series. To this end, I use a Hodrick-Prescott filter to capture slow moving conditional mean process. I also experimented with the method suggested by Bansal and Yaron (2004) and Bansal, Kiku, and Yaron (2007) which regress consumption growth onto interest rate and price-dividend ratio, exploiting the theoretical structure. Although both results are compatible with each other, I find that statistical filtering seems to capture long-run components better, especially with monthly frequency data.

Now with $y_{j,t}$ in hand, one can easily see that (3.13) as well as (3.5) to (3.8) form a state space model. Thus, a simple method in the conventional approach to this model could be an extended Kalman filtering by linearizing (3.13) after taking squares and logs. However, the extended Kalman filtering usually generates the bias in the estimation. The linearization and the normality assumption on error terms are the main source of the bias. In the persistent latent variable case, additionally, the integrability of the derivative of the logistic function makes Kalman gain disappear at the near of lower bound or upper bound. See the chapter IV. An alternative method is the density-based nonlinear filtering using exact density function during prediction

steps and updating steps. I derive nonlinear filtering algorithm to this model in Appendix D. However, this algorithm is hard to implement in terms of speed due to multi dimensional integration. So I take the the Bayesian approach, in particular, the Gibbs sampling which does not suffer seriously the curse of dimensionality because it utilizes univariate conditional density function. While the Gibbs Sampling method was originally introduced for image restoration by Geman and Geman (1984), it has widely been used to solve state-space models, in particular, SV models. See the Jacquier, Polson, and Rossi (1994), So, Lam, and Li (1998), Jacquier, Polson, and Rossi (2004), and Geweke and Tanizaki (2001).

I view both the latent variables and the parameters as random variables generated from the joint posterior density function $p(L, \Psi|Y)$, where $Y = (y_1, \cdots, y_n)$, $y_t = (y_{c,t}, y_{d,t})$, $L = (X, W)$, $X = (x_1, \cdots, x_n)$, $x_t = (x_{c,t}, x_{d,t})$, $W = (w_1, \cdots, w_n)$ and $\Psi = (\theta, \rho, \lambda, \sigma^2)$. This augmentation of latent variable has originally been introduced by the Tanner and Wong (1987).[3] As discussed in Jacquier, Polson, and Rossi (1994), the fact that $p(L, \Psi|Y)$ is proportional to conditional density of observable variables $p(Y|L, \Psi)$, conditional density of unobservable variables $p(L|\Psi)$, and prior density of parameters $p(\Psi)$ shows that the joint posterior captures the hierarchical structure of this model well. Our goal is to sample (L, Ψ) from the joint posterior distribution. Once I draw the observations for (L, Ψ), I actually have samples for the parameters generated by $p(\Psi|Y) = \int p(L, \Psi|Y)\, dL$ and samples for the latent variables generated by $p(L|Y) = \int p(L, \Psi|Y)\, d\Psi$. Then I can get parameter estimation by the Monte-

[3]The conventional approach tries to find Ψ maximizing likelihood function $p(\Psi|Y)$. And then it uses $p(L|Y, \hat{\Psi})$ to get updated latent variables $L_{t|t}$ or smoothed latent variables $L_{t|n}$.

Carlo integration i.e.,

$$E[\Psi|Y] = \int \Psi \; p(\Psi|Y)\, d\Psi \approx \frac{1}{G} \sum \Psi^{(i)}$$

where $\Psi^{(i)}$ is the i-th sample for Ψ and G is the total number of valid Gibbs samples for Ψ. And similarly I can get the smoothed latent variables. In this model, I can derive the joint posterior distribution as following:

$$\begin{aligned} p(L, \Psi|Y) &\propto p(L, Y|\Psi)p(\Psi) \\ &\propto \left(\prod_{t=1}^{n} p(y_t|x_t, \Psi)p(x_{c,t}|w_t, \Psi)p(x_{d,t}|w_t, \Psi)p(w_t|w_{t-1}) \right) p(\Psi) \end{aligned} \tag{3.14}$$

Though I know the joint posterior distribution it is not easy to draw samples from it directly. So the Gibbs sampling constructs a Markov chain of (L, Ψ) where each step draws sample of L or Ψ from a standard posterior density conditional on all the available information. For example, draw a latent variable from $p(L|Y, \Psi)$ and draw a parameter from $p(\Psi|Y, L)$ back and forth. The beauty of the Gibbs sampling is that it is asymptotically identical to the sampling directly from the joint posterior distribution under mild conditions[4] as proved by Tierney (1994). So our next step is to find posterior density function of each random variables and draw a sample from it.

First, I can derive easily the posterior distribution of each latent variable conditional on all the other information from (3.14).[5] See the Appendix E for details. For

[4]Mild conditions are irreducibility and aperiodicity which are satisfied by our transition kernels.

[5]By Carter and Kohn (1994) block sampler which samples X simultaneously is more efficient than the sampler that generates the latent variables one at a time when state space model is linear. But I cannot use block sampler here because of nonlinearity.

common latent factor w_t,

$$p(w_t|X, W_{\setminus t}, Y, \Psi) \propto p(x_{c,t}|w_t, \lambda_c, \sigma_c^2)p(x_{d,t}|w_t, \lambda_d, \sigma_d^2)p(w_{t+1}|w_t)p(w_t|w_{t-1}) \tag{3.15}$$

where $W_{\setminus t}$ denotes the rest of the W vector other than w_t. (3.15) shows that $w_t|X, W_{\setminus t}, Y, \Psi$ follows a normal distribution of $\mathbb{N}\left(B_w A_w^{-1}, A_w^{-1}\right)$ where $A_w = \sum_j \frac{\lambda_j^2}{\sigma_j^2} + 2$, $B_w = \sum_j \frac{\lambda_j x_{j,t}}{\sigma_j^2} + w_{t+1} + w_{t-1}$.[6] So it is easy to sample a common latent variable from (3.15). But this is not the case for $x_{j,t}$. The posterior density function for $x_{j,t}$ can be derived as following:

$$p(x_{j,t}|X_{\setminus j,t}, W, Y, \Psi) \propto p(y_t|x_t, \theta, \rho)p(x_{j,t}|w_t, \lambda_j, \sigma_j^2) \tag{3.16}$$

The complexity of (3.16) makes it impossible to draw a sample from the posterior distribution of $x_{j,t}|X_{\setminus j,t}, W, Y, \Psi$ directly. The Metropolis-Hastings (MH) algorithm is useful for such complicated distribution. MH algorithm draws a sample from a transition kernel chosen by researcher. If this transition kernel is ideal, the distribution of the samples generated from the transition kernel is approximately the target distribution. Since the ideal transition kernel is unknown to the researcher, MH algorithm compares the probability of the new sample with that of previous sample. If the former is higher, MH algorithm accepts the new sample as a realization from the target distribution for sure. But if the latter is higer, MH algorithm accepts the new sample as a realization from the target distribution with some probability[7] such

[6] for $t = n$, $A_w = \sum_j \frac{\lambda_j^2}{\sigma_j^2} + 1$ and $B_w = \sum_j \frac{\lambda_j x_{j,n}}{\sigma_j^2} + w_{n-1}$

[7] The probability of move is

$$\alpha(x, y) = min\left[\frac{p(y)q(y, x)}{p(x)q(x, y)}, 1\right]$$

where x is current location and y is next location, p is the target density, $q(x, y)$ is a transition kernel which generates a value y when the process is at the point x.

that nth iterate of the transition kernel converges to the target distribution. When MH algorithm rejects the new sample, it accepts the previous sample as the new realization. The reader is referred to Chib and Greenberg (1995) for more detail MH algorithm. Geweke and Tanizaki (2001) shows that when I use MH algorithm to draw the latent variables the most efficient transition kernel is a density function obtained from the transition equation. So I choose $\frac{1}{\sqrt{2\pi\sigma_j^2}}\exp\left[-\frac{(x_{j,t}-\lambda_j w_t)^2}{2\sigma_j^2}\right]$ as transition kernel for latent variables $x_{j,t}$. This kernel is independent of the previous realization.

Next, I move on the posterior distribution of parameters. Here I need their prior distributions. I assume the α, β, σ^2 and λ have Gamma distribution due to the positiveness and κ and ρ follow normal distribution. Then the posterior distributions can be derived as following: for $\Psi_1 = (\theta, \rho)'$

$$p(\Psi_1|X, Y, W, \Psi_{\setminus 1}) \propto \left(\prod_{t=1}^{n} p(y_t|x_t, \Psi_1)\right) p(\Psi_1) \tag{3.17}$$

and for $\Psi_2 = (\lambda, \sigma^2)'$,

$$p(\Psi_2|X, W, Y, \Psi_{\setminus 2}) \propto \left(\prod_{t=1}^{n} p(x_t|w_t, \Psi_2)\right) p(\Psi_2). \tag{3.18}$$

Since (3.17) and (3.18) are already so complicated, I need to rely on MH algorithm again. Following the argument of Geweke and Tanizaki (2001) I choose the prior distribution as the transition kernel.

2. Data and Gibbs Sampling Results

I use monthly consumption and dividend series from February 1959 to December 2006 for applying the method I developed. Consumption, nondurable goods and service expenditure series are obtained from Fed St. Louis web site[8] and dividend data from

[8]http://research.stlouisfed.org.

Table 10.: Summary Statistics.

	Mean	Max	Min	Std.Dev	Skewness	Kurtosis
Consumption	0.0000	0.0119	-0.0144	0.0035	-0.1174	4.0933
Dividend	0.0000	0.0743	-0.1370	0.0312	-0.9544	4.7895

Table 11.: Gibbs Sampling Results.

		Posterior		Convergence
Parameters	Priors	Mean	Standard errors	Diagnostics
α_c	$G(1,2)$	0.000003	0.000001	-1.526
β_c	$G(1,4)$	0.000023	0.0000089	3.241
κ_c	$N(0,1)$	-0.0554	0.9391	3.415
λ_c	$N(0.5,1)$	0.2277	0.1355	-2.840
σ_c^2	$G(1,2)$	1.7174	0.6204	-2.225
α_d	$G(1,3)$	0.000142	0.0000408	6.293
β_d	$G(5,10)$	0.003522	0.0009211	0.617
κ_d	$G(0,1)$	2.0719	0.7107	-1.697
λ_d	$G(0.5,1)$	0.3605	0.1041	4.778
σ_d^2	$G(1,2)$	0.4112	0.2118	2.478
ρ	$N(0,1)$	-0.0258	0.0436	-0.694

Notes: $G(a,b)$ denotes gamma distribution with mean ab and variance ab^2 and $N(a,b^2)$ denotes normal distribution with mean a and variance b^2.

CRSP. Dividend growth data is generated from seasonally-adjusted real dividend series which is created using value-weighted returns with dividend and without dividend.[9] There are altogether 575 observations. Table 10 provides summary statistics for two series. As discussed earlier, I use hp-filter to subtract predictable component of each series and I demean hp-filtered series to make conditional mean equal to zero.

I draw 120000 samples for each parameters and latent variables by Gibbs algo-

[9]Real dividend = (returns with dividend - returns without dividend) * market price index / CPI.

rithm and discard first 40000 samples, which are considered as samples in the burn-in period. Table 11 shows most of the estimated parameters are well converged and significant. Though β_c and α_d have relatively high convergence diagnostics[10], their small standard errors show that sample series are quite concentrated after burn-in period as illustrated in Figure 13 and Figure 14. Jacquier, Polson, and Rossi (2004) warns that one must perform careful sampling experiments to establish convergence across a wide range of empirically relevant parameter values. Our samples show that they wander a lot at the beginning but converge after burn-in period.

I can see the α_c and β_c is smaller than α_d and β_d due to the volatile dividend process. While lower bound of volatility of consumption and dividend are $\sqrt{\hat{\alpha}_c} = 0.02\%$ and $\sqrt{\hat{\alpha}_d} = 1.19\%$, upperbound of volatility function of consumption growth and dividend are $\sqrt{\hat{\alpha}_c + \hat{\beta}_c} = 0.49\%$ and $\sqrt{\hat{\alpha}_d + \hat{\beta}_d} = 6.05\%$ in a month respectively. $\hat{\lambda}_c$ and $\hat{\lambda}_d$ show each volatility generating process scale down the magnitude of macro uncertainty. $\hat{\sigma}_c^2$ is larger than $\hat{\sigma}_d^2$ which implies that the signal of dividend process is stronger than that of consumption process so that dividend volatility generating process ($x_{d,t}$) is more similar to common macro volatility generating process (w_t). The difference of all these parameters indicates that the assumption of Bansal and Yaron (2004) that conditional volatility of dividend is a simple constant multiplication of

[10]Convergence Diagnostic (CD) is given by

$$CD = \frac{\overline{\theta}_A - \overline{\theta}_B}{\sqrt{\frac{\hat{f}_A(0)}{n_A} + \frac{\hat{f}_B(0)}{n_B}}}$$

where A is the set of Gibbs samples with n_A iterations after burn-in period and B is the set of Gibbs samples with last n_B observations and f(0) is the spectral density at the zero frequency with Parzen window. By the convention I set $n_A/n = 0.1$ and $n_B/n = 0.5$ where n denotes the number of Gibbs samples after burn-in periods. Geweke (1992) proves that if the sequence of Gibbs samples for a parameter is stationary, CD converges to the standard normal distribution as the number of samples goes to the infinity.

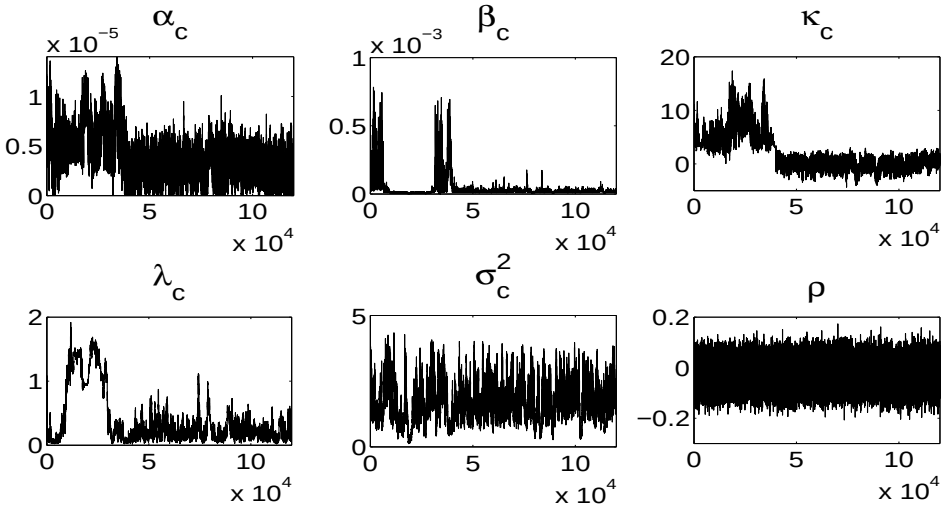

Fig. 13.: Gibbs Samples for Parameters of Consumption Growth.

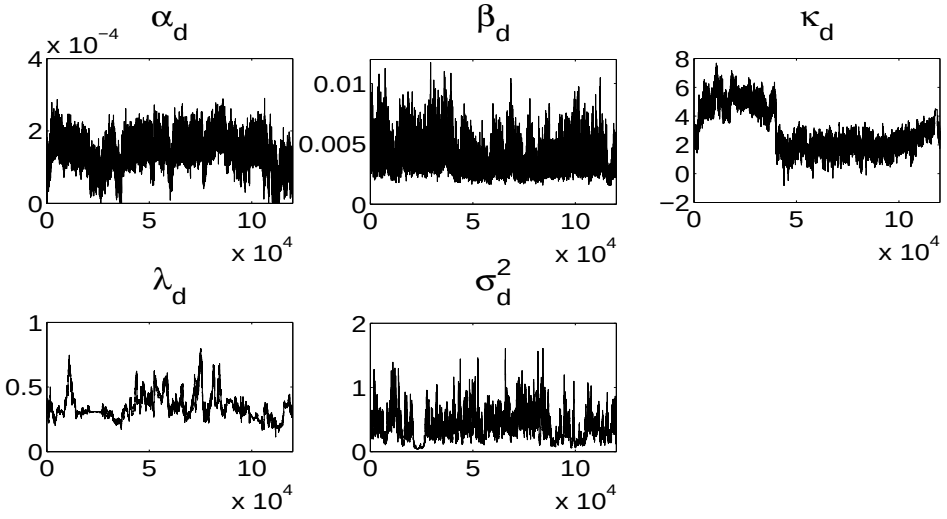

Fig. 14.: Gibbs Samples for Parameters of Dividend Growth.

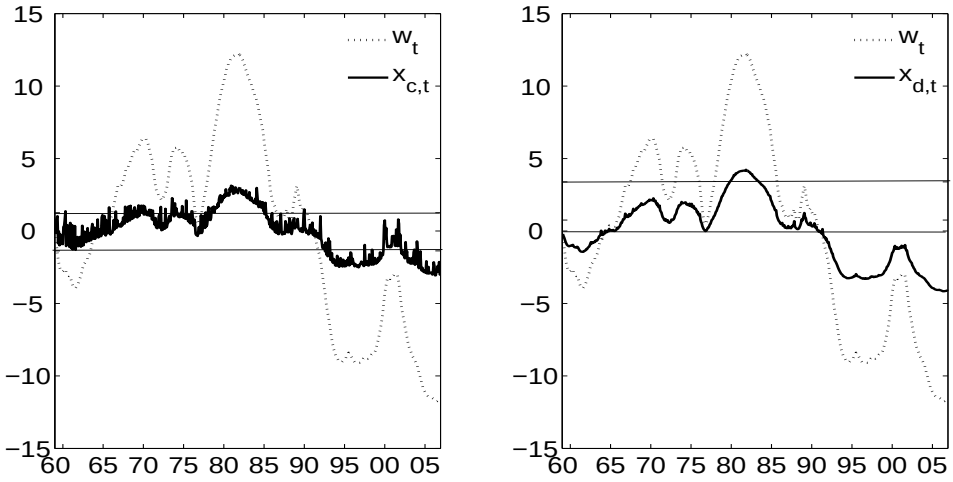

Fig. 15.: Comparison between Volatility Generating Process.

conditional volatility of consumption may not be correct, though they have a common features. $\hat{\rho}$ shows that the correlation between short-run shocks is not big.

I also extract the latent common macroeconomic uncertainty process w_t and volatility generating process for consumption and dividend $(x_{c,t})$, $(x_{d,t})$ as illustrated in Figure 15. All three volatility generating process have unit root and volatility generating process of consumption and dividend are cointegrated with common volatility generating process.[11] Straight line can be thought as a threshold for the regime change in a rough sense.[12] The area above the upper straight line is high volatility period and vice versa. So I can interpret that the fundamentals which generates volatility have stayed in the low volatility regime in 60s and moved into the transition period in

[11]ADF unit root test t-statistics are -1.2033, -1.2328 and -1.5258 respectively while the critical value for 5 % confidence level is -2.866. Residual based cointegration test t-statistics between w_t and $x_{c,t}$ are -20.9900 and that of between w_t and $x_{d,t}$ are -4.0182 while the critical value for 5 % confidence level is -3.46.

[12]I may think X $= [\kappa - \frac{\log(2+\sqrt(3))}{\lambda},\ \kappa - \frac{\log(2-\sqrt(3))}{\lambda}]$ as transition period because $\frac{\partial^3 f(x)}{\partial x^3} = 0$ at $x = \kappa - \frac{\log(2\pm\sqrt(3))}{\lambda}$.

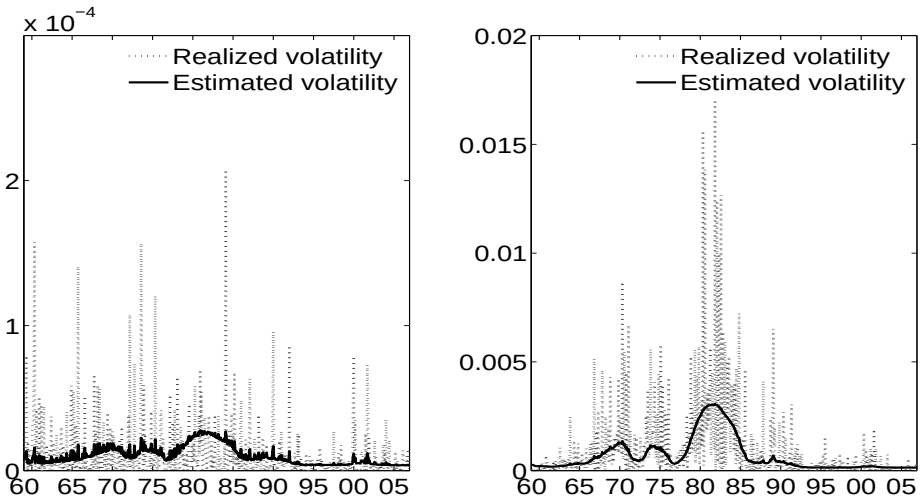

Fig. 16.: Realized and Estimated Volatility of Consumption and Dividend .

70s and entered into high volatility regime in first-half 80s. From late 80s until now the fundamentals have stayed in the transition period or the low volatility regime.

Figure 16 plots the estimated volatility and realized volatility. Estimated volatility from Gibbs samples is obtained by Monte-Carlo integration i.e.,

$$\begin{aligned} E[f_j(x_{j,t})|\mathcal{F}_n] &= \int f_j(x_{j,t}) p(x_{j,t}|\mathcal{F}_n)\, dx_{j,t} \\ &\approx \frac{1}{G} \sum_{i=k+1}^{k+G} f_j(x_{j,t}^{(i)}) \qquad \text{for} \quad j = c, d \end{aligned}$$

where $x_{j,t}^{(i)}$ is the i-th sample for $x_{j,t}$ and k is the length of burn-in period and G is the total number of Gibbs samples for $x_{j,t}$. I see from Figure 16 that generally the estimated volatilities explains well the realized volatilities.

Finally, I try to match the macro uncertainty process with some macro-economic fundamentals. Since the purpose of this paper is not the interpretation of the volatility generating process in a serious manner I document just the relationship that could give a meaningful interpretation. Figure 17 shows the relationship between monthly

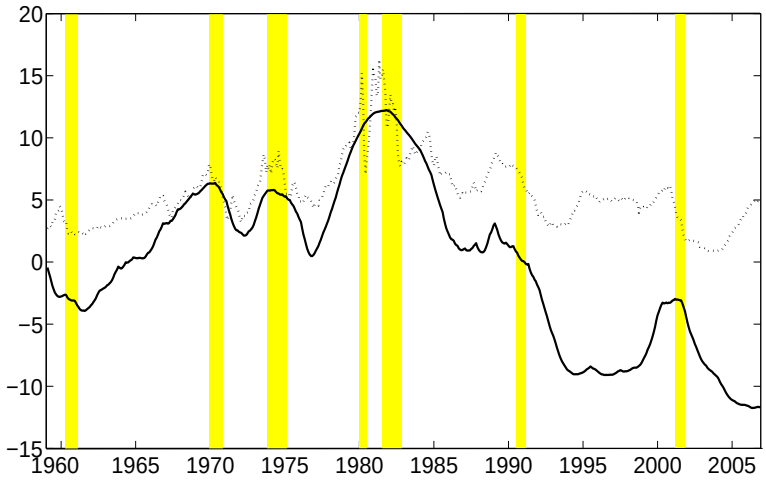

Fig. 17.: Nominal Interest Rates and Macro Uncertainty.

Notes: The solid line stands for the common volatility generating process and the dotted line stands for nominal interest rates of 3 month Treasury bill and shadow area stands for the recession period.

nominal interest rates and the common volatility generating process. The interest rates data is obtained from 3-month treasury bond at St. Louis Fed. Monthly nominal interest rates and the volatility generating process seems to share the common stochastic trend.[13] And I find that recession period are well matched with peak of common volatility generating process. Especially common volatility generating process reaches its peak just before the entry into the 70s, 80s and recent 2001 recession, and even 1991 mini credit crunch is predicted by a small, but conspicuous increase in the macro uncertainty factor w_t. One of the most foundational links between asset returns and macroeconomic variables is that equity premium is higher at business cycle troughs, reflecting risk-return tradeoffs with counter cyclical variations. Our

[13]Cointegration test for whole series cannot reject null hypothesis. But when I narrow the period between 1959 and 1990, the Johansen Cointegration Test confirms the cointegration relationship. p-value with Maximum Eigenvalue is 0.0419 and with Trace 0.0097.

macroeconomic uncertainty factor rises and becomes highest when recession starts, i.e. time of highest uncertainty. Thus, I can expect that this is going to be well connected to higher risk premium demanded by asset market participants. In the next section, I estimate this relationship using the identified state variables.

D. Equity Premium: Estimation Results

I use Generalized Moment Method (GMM) to estimate two core preference parameter, risk aversion coefficient γ and intertemporal elasticity of substitution ψ as Bansal, Kiku, and Yaron (2007). GMM use only the moment condition derived from Euler equation (3.1) without an additional assumption on the distribution. I use the GMM code downloaded from the website.[14] I use moment conditions for market, large, small, growth and value stock and one period lagged consumption growth as instrument. Each asset data is obtained from Fama and French data library [15] and risk-free rate is 3 month Treasury Bill rate obtained from St. Louis Fed and they are all in monthly frequency and real terms.

I find γ is 2.111 and ψ is 2.115. Mehra and Prescott (1985) consider that a maximal level for risk aversion is around 10. And Barro (2005) argues that the usual view in the finance literature is between 2 and 5. In this sense, our finding for risk aversion is very reasonable. And several papers consider that plausible ψ lies in between 1 and 2. For example, Bansal and Yaron (2004) benchmark value is 1.5 and Hansen and Singleton (1982) and Attanasio and Weber (1989) and Vissing-Jorgensen (2002) estimate ψ to be well in excess of 1.5 and 1 respectively. In most studies, however, their measurements are noisy and our case is not an exception. At least, I

[14] http://www.mgmt.purdue.edu/faculty/mcliff/progs.html.

[15] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data-library.html.

Table 12.: GMM Estimation Result and Moment Conditions.

GMM Parameter Estimates				
Parameters	Estimates	Standard errors	t-stat	p-val
γ	2.1112	0.3798	5.56	0.0000
ψ	2.1153	41.5402	0.05	0.9594
	GMM Moment Conditions			
	Moment	Standard errors	t-stat	p-val
Moment 1	0.003152	0.002374	1.33	0.1843
Moment 2	0.001702	0.00106	1.61	0.1083
Moment 3	-0.001216	0.000678	-1.79	0.073
Moment 4	-0.00011	0.000262	-0.42	0.6753
Moment 5	-0.00154	0.001443	-1.07	0.2859
Moment 6	-0.000419	0.000541	-0.77	0.4385
Moment 7	0.004064	0.001317	3.09	0.002
Moment 8	0.002092	0.000778	2.69	0.0072
Moment 9	-0.000613	0.000428	-1.43	0.1515
Moment 10	-0.000029	0.000275	-0.10	0.9165
	J-start = 15.81	$\text{Prob}(\chi 2_{(8)} > J) = 0.045$		

Notes: Moment condition 1-5 are Return on Small, Large, Growth, Value and Market. And Moment condition 6-10 are orthogonality between each returns and one period lagged consumption growth. Since I have moment conditions more than parameters, I use identity matrix as starting weighting matrix and generate optimal weighting matrix.

Table 13.: Model Implied Asset Returns.

	γ	ψ	$E(r_f)$	$\sigma(r_f)$	$E(r_m)$	$\sigma(r_m)$	$E(r_m - r_f)$
Baseline	2.111	2.115	0.0187	0.0142	0.0880	0.1923	0.0693
Data			0.0351	0.0189	0.0860	0.1519	0.0509
Changes in γ	2.111	2.115	0.0187	0.0142	0.0880	0.1923	0.0693
	3		0.0138	0.0142	0.1107	0.2853	0.0968
	5		0.0043	0.0142	0.2073	0.4783	0.2030
	10		-0.0142	0.0142	0.7070	0.9040	0.7212
Changes in ψ	2.111	1.5	0.0227	0.0199	0.0888	0.1732	0.0662
		2.115	0.0187	0.0142	0.0880	0.1923	0.0693
		2.5	0.0170	0.0120	0.0879	0.1996	0.0710
		3	0.0154	0.0100	0.0888	0.2063	0.0734

confirm that our finding for IES is not far away from consensus in finance literature.

As discussed in the previous subsection, I simulate the 1000 data points series of consumption and dividend 1000 times. I use GMM estimation results and its by-product, all the constant terms to be used in generating process. The standard deviations of simulated consumption and dividend process are 0.0045 and 0.0222 respectively. Since the those of actual data are 0.0036 and 0.0324, our simulated data is plausible well. From these simulated data I have asset returns which is reported as the baseline in Table 13. Our simulated equity premium is 6.93 % and risk free rate is around 2 %. This is consistent with most of the empirical results and I can match the second moments of equity premium and risk free rate as well. Even compared with the existing calibration and empirical literature, our model seems to perform quite well with both the risk aversion and IES around two. I conjecture that the results come from more realistic modeling strategy of macroeconomic uncertainty.

I run the sensitivity test to understand the role of risk aversion coefficient and IES. First I increase the risk aversion coefficient from the baseline. I see that very

common phenomenon that more risk averse preference generates higher equity premium, too far from the actual data. Secondly I changes in IES. I vary IES between 1.5 and 3 and its effect is as I expect from the derived equations for risk free rate and excess returns. But, differences are small.

E. Conclusion

In this paper I try to link the macroeconomic uncertainty and asset pricing because macroeconomic uncertainty, which is time-varying but unobservable, is considered as an important ingredient for asset pricing. First, I introduce a stochastic volatility model with consumption and dividend process to identify macroeconomic uncertainty. In order to capture the features of volatility in financial and macroeconomic data, I assume the volatility function takes logistic form and unit root common factor. I solve this model numerically with Bayesian approach to avoid the multidimensional difficulties. I find that the extracted volatility series explains well the realized volatility series of both consumption and dividend. Also I see a counter-cyclical relation of the extracted macroeconomic uncertainty.

And then motivated from Bansal and Yaron (2004), I combine this stochastic volatility model with long-run risk model and Epstein-Zin-Weil preference. I find our estimated risk-aversion coefficient and intertemporal elasticity of substitution around two which is plausible according to the consensus in the finance literature. Bansal and Yaron's model with relatively high risk aversion can generate high risk premium through the persistent long-run risk channel. However, our model produces high risk premium even with moderate coefficient of risk aversion because it has another channel, more realistic time-varying volatility. Furthermore, I find that the market return is as volatile as real data and the risk-free rate is low and stable.

CHAPTER IV

NONLINEAR FILTERING WITH A LATENT AUTOREGRESSIVE STATE: A NUMERICAL COMPARISON OF FOUR TECHNIQUES

A. Introduction

Filtering of an unobserved state is a well-researched topic of interest to physicists, engineers, econometricians, and other researchers across a wide range of fields of inquiry. The most well-known approach, developed by Kalman (1960) was designed to handle linear state space systems. The state is assumed to be autoregressive, and the observation of the state is assumed in such systems to be a simple linear transformation of the state, augmented with noise.

Although linearity is a cornerstone of the Kalman filter (KF), engineers quickly found the need to model nonlinear phenomena, such as tracking the trajectory of an aircraft, spacecraft, etc. Early attempts to handle nonlinearity included the extended Kalman filter (EKF) (Jazwinski (1970)), which uses linear approximations of the nonlinearities in the observation and/or state equation. The EKF introduces bias in both the linear approximation(s) of the nonlinear function(s) and also in the use of linear updating of predictions of the state.

Alternative techniques aimed at mitigating the inherent biases of the EKF include techniques that sample – either randomly or nonrandomly – from an untransformed distribution. Once propagated through the transformation, the resulting points correspond to a transformed distribution that yields information about the transformation itself. The unknown parameters of this transformation may thus be estimated.

In this analysis, I compare such approaches for a range of prototypical nonlinear state space models with a single autoregressive state. Specifically, I compare the KF

and EKF to a technique that uses nonrandom sampling – the unscented Kalman filter (UKF) (Julier and Uhlmann (1997) ;Julier, Uhlmann, and Durrant-Whyte (1993)) – and to a technique that uses random sampling – the density-based nonlinear filter (DNF) (Harvey and Shepard (1996);Tanizaki (1996)).

As econometricians, an autoregressive state is particularly appealing, since many time series are assumed to follow I(0) or I(1) stochastic processes in the economic and econometric literatures. Our focus on a univariate state precludes the potential complication of cointegrated states in a multivariate setting. Cointegration would substantially alter the statistical properties of all series considered and is beyond the scope of the present analysis.

For linear models, the asymptotic distributions of parameter estimators and a closed-form relationship between the latent series and the series of conditional estimates is complicated but tractable. (See, e.g., Chang, Miller, and Park (2006).) For nonlinear models, iterative prediction and updating create infinite-order, stochastic, nonlinear difference equations. Asymptotic distributions of the parameter estimators are generally intractable, and closed-form expressions are scarce. In this light, I rely on simulations to gain insights into the relative advantages in finite samples of each of the fours filters under alternative nonlinear specifications. I examine four aspects of estimator performance: parameter estimation, state estimation, in-sample fit of the model, and numerical stability of the algorithm.

Our finite sample results offer some interesting and unexpected insights. Clearly, the linear KF cannot estimate the nonlinear parameters of the true model, and the state is not estimated well. However, the overall in-sample fit and numerical stability of this linear technique is surprisingly good for our nonlinear models. The three nonlinear techniques can estimate the parameters of a well-specified model, with some performing better with some functions and worse for others. The EKF seems to be

the best with in-sample fit of the model, and it is the most numerically stable of the three for most functions. However, one or both of the other two nonlinear techniques (UKF and DNF) usually outperform the EKF at parameter and state estimation. The DNF and UKF filters appear to work well for estimating these, although the DNF has weaknesses when the model has parameters that are not well-identified. On the other hand, these two are generally the least numerically stable algorithms.

Econometricians have been increasingly interested in nonlinear filtering over the last two decades, as more and more nonlinear structural modeling applications with latent states or time-varying parameters have presented themselves. The tool of choice among econometricians is by far the EKF. To highlight the apparent appeal of this technique – and the implicit demand for better nonlinear techniques – I briefly review some macroeconometric and financial econometric analyses that have relied on the EKF.

The EKF has been used in the econometrics literature to filter series through a nonlinear function, similarly to the class of models I discuss below. For example, Bijleveld, Commandeur, Koopman, and van Montfort (2007) used the EKF to model road safety with exponential risk factors. Miller and Park (2008) used it to estimate target zone exchange rate models with a latent economic fundamental.

This technique is also a popular technique for estimating macroeconomic models with time-varying parameters. Grillenzoni (1993) suggested using the EKF to estimate ARIMA models with ARIMA parameters. Simplification of such models may be estimated using a linear filter. However, models that feature both time-varying parameters and a latent state necessarily require a nonlinear technique. Bacchetta and Gerlach (1997) employed such an approach to examine a consumption function with a time-varying proportion of consumers with credit constraints.

Moreover, this technique is practical for models with inherently nonlinear func-

tions and that also require filtering of time-varying parameters. For example, Tanizaki (1993) introduced the EKF for binary choice models with time-varying parameters. Shen, Hakes, and Brown (1999) applied this technique to estimate a binary monetary reaction function for the Fed.

The EKF is also well-known in finance. Bolland and Connor (1997) used the EKF when the state is assumed to have a nonlinear dynamic (artificial neural network) structure. A number of papers (Lund (1994); Duan and Simonato (1997); Duffee (1999); Geyer and Pichler (1999); Chen and Scott (2003); An (2007)) have used the EKF to estimate term structure models.

As interest in the EKF continues to grow among econometricians, our numerical comparison with alternative nonlinear filters provides timely insights into the relative strengths and weaknesses of this filter.

The remainder of the paper is structured as follows. In the following section, I outline a prototypical nonlinear state space model with univariate autoregressive state. I then discuss in more detail the four filtering technique mentioned above: KF, EKF, UKF, and DNF. I detail our simulation experiments in Section C, and I present and discuss our numerical results in Section D.

B. Nonlinear Filtering in Theory and Practice

A general nonlinear filtering problem involves observation (or measurement) and state (or transition) equations featuring nonlinear functions with multivariate arguments. I consider a single, univariate state with linear autoregressive structure. Specifically, our observation and state equations are given by

$$\begin{aligned} y_t &= g(x_t, \theta) + u_t \\ x_t &= \rho x_{t-1} + v_t \end{aligned}$$

where

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim \mathbf{iidN}\left(0, \begin{pmatrix} \sigma^2 & 0 \\ 0 & 1 \end{pmatrix}\right),$$

and (y_t) and (x_t) are scalar series with latent (x_t). The restriction of the variance of (v_t) is necessary for parameter identification under some specifications. The function g is known up to an unknown parameter vector θ. The initial condition x_0 is assumed to be independent of (u_t) and (v_t).

This framework may easily be generalized to take into account serial correlation in (v_t) by allowing (x_t) to be multivariate. However, as discussed above, I do not consider this case in our investigation in order to eschew complications from cointegration.

For a known parameter vector (ρ, θ, σ^2), estimation of the latent series (x_t) is accomplished by estimating a series of conditional expectation of the latent series given information made available by the observed series (y_t). For an unknown parameter vector, I can implement such a filter within a numerical optimization procedure, so that the resulting filter uses parameter estimates that maximize the joint log-likelihood of (y_t). All of our simulations involve such numerical optimization.

Filtering generally proceeds in three steps: [P] one-step-ahead prediction of the conditional density of x_t, [L] conditional likelihood calculation, and [U] updating the prediction with newly available information. If I define an increasing filtration $(\mathcal{F}_t)$ as the natural filtration for (y_t) – *not including* (x_t) – then new information at time t is given by $y_t - y_t|\mathcal{F}_{t-1}$. A general outline of these steps follows:

[P] **Prediction.** One-step-ahead prediction of the conditional density $p\left(x_t|\mathcal{F}_{t-1}\right)$ of the state is given by

$$p\left(x_t|\mathcal{F}_{t-1}\right) = \int p\left(x_t|x_{t-1}\right) p\left(x_{t-1}|\mathcal{F}_{t-1}\right) dx_{t-1}$$

using the assumption that (v_t) is contemporaneously independent of (y_{t-1}). In

this model, I may more specifically write this as

$$p\left(x_t|\mathcal{F}_{t-1}\right) = p\left(\rho x_{t-1}|\mathcal{F}_{t-1}\right) + p\left(v_t\right), \tag{4.1}$$

due to the linearity of the state equation and the normality of (v_t).

[L] **Likelihood calculation.** The conditional density $p\left(y_t|\mathcal{F}_{t-1}\right)$ is given by

$$p(y_t|\mathcal{F}_{t-1}) = \int p(y_t|x_t)p(x_t|\mathcal{F}_{t-1})dx_t$$

using the assumption that (u_t) is contemporaneously independent of (x_t). I may more specifically write this as

$$p(y_t|\mathcal{F}_{t-1}) = p(g\left(x_t,\theta\right)|\mathcal{F}_{t-1}) + p\left(u_t\right), \tag{4.2}$$

due to the linear separability of $g\left(x_t,\theta\right)$ and u_t in the observation equation and the normality of (u_t).

[U] **Updating.** In order to update the prediction with information available at time t and to subsequently predict to time $t+1$, I require the conditional density $p\left(x_t|\mathcal{F}_t\right)$. This is generally obtained from

$$p\left(x_t|\mathcal{F}_t\right) = p(x_t|y_t,\mathcal{F}_{t-1}) = \frac{p(y_t|x_t)p(x_t|\mathcal{F}_{t-1})}{p(y_t|\mathcal{F}_{t-1})}, \tag{4.3}$$

using, again, the assumption that (u_t) is contemporaneously independent of (x_t).

This filtering strategy may be accomplished iteratively, subject to a known parameter vector θ, a known density of the initial state $p\left(x_0|\mathcal{F}_0\right)$, and a known (normal) likelihood $p(y_t|x_t)$.

Throughout the remainder of the paper, I adopt the following notation from the

literature. I let

$$y_{t|\bullet} \equiv \mathbf{E}y_t|\mathcal{F}_\bullet \qquad \text{and} \qquad x_{t|\bullet} \equiv \mathbf{E}x_t|\mathcal{F}_\bullet$$

denote conditional means, and

$$\omega_{t|\bullet} \equiv \mathbf{var}(x_t|\mathcal{F}_\bullet), \qquad \sigma_{t|\bullet} \equiv \mathbf{var}(y_t|\mathcal{F}_\bullet), \qquad \text{and} \qquad \xi_{t|\bullet} \equiv \mathbf{cov}(x_t, y_t|\mathcal{F}_\bullet)$$

denote conditional variances and covariance.

1. Kalman Filter

In order to motivate the nonlinear filters discussed below and in order to provide a benchmark technique, I briefly discuss the Kalman filter. When the observation equation is linear, say,

$$y_t = \alpha + \beta x_t + u_t$$

the Kalman filter is the most widely used technique for estimating (x_t).

[P] Prediction is accomplished from (4.1) by

$$\begin{aligned} x_{t|t-1} &= \rho x_{t-1|t-1} \\ \omega_{t|t-1} &= \rho^2 \omega_{t-1|t-1} + 1 \end{aligned}$$

since $(x_{t-1|t-1})$ and (v_t) are normal.

[L] Similarly, the conditional likelihood is calculated from (4.2) using

$$\begin{aligned} y_{t|t-1} &= \alpha + \beta x_{t|t-1} \\ \sigma_{t|t-1} &= \beta^2 \omega_{t|t-1} + \sigma^2 \end{aligned}$$

since $(x_{t|t-1})$ and (u_t) are normal.

[U] Slightly more complicated, updating is given by

$$x_{t|t} = x_{t|t-1} + K_{t|t-1}\left(y_t - y_{t|t-1}\right) \tag{4.4}$$
$$\omega_{t|t} = \omega_{t|t-1} - K_{t|t-1}^2\sigma_{t|t-1}$$

where

$$K_{t|t-1} \equiv \xi_{t|t-1}\sigma_{t|t-1}^{-1} = \beta\omega_{t|t-1}\left(\beta^2\omega_{t|t-1} + \sigma^2\right)^{-1}$$

is the so-called Kalman gain. Note that $K_{t|t-1}\left(y_t - y_{t|t-1}\right)$ is simply an orthogonal projection of $x_t - x_{t|t-1}$ onto the space of $y_t - y_{t|t-1}$, which represents new information available at time t. The linear projection optimally estimates $p(x_t|y_t, \mathcal{F}_{t-1})$, since the model is linear.

2. Extended Kalman Filter

The EKF (Jazwinski (1970)) was popularized in the statistical and econometric literatures by Harvey (1990). Assuming differentiability of g, the EKF relies on a simple Taylor-type expansion of the nonlinear function $g\left(x_t\right)$ around the "known" but unobservable series $(x_{t|t-1})$, such that

$$g\left(x_t\right) = g\left(x_{t|t-1}\right) + G_{t|t-1}\left(x_t - x_{t|t-1}\right) + r_t$$

where $G_\bullet$ denotes the first derivative of $g\left(x_t\right)$ evaluated at $x_t = x_\bullet$ and r_t represents all remainder terms.

[P] Prediction is the same as for the Kalman filter, since I assume a linear state equation.

[L] In order to evaluate the likelihood given by (4.2), I need to evaluate $\mathbf{E}g(x_t)|\mathcal{F}_{t-1}$

and $\mathbf{var}\left(g(x_t)|\mathcal{F}_{t-1}\right)$. Using the above linearization, I have

$$\begin{aligned}
\mathbf{E}g(x_t)|\mathcal{F}_{t-1} &= g\left(x_{t|t-1}\right) + r_{t|t-1} \\
\mathbf{var}\left(g(x_t)|\mathcal{F}_{t-1}\right) &= G_{t|t-1}^2\omega_{t|t-1} + \mathbf{var}\left(r_t|\mathcal{F}_{t-1}\right) + \mathbf{cov}\left(G_{t|t-1}\left(x_t - x_{t|t-1}\right), r_t|\mathcal{F}_{t-1}\right)
\end{aligned}$$

The first-order EKF approximates r_t with zero, so that (4.2) is characterized by

$$\begin{aligned}
y_{t|t-1} &\approx g\left(x_{t|t-1}\right) \\
\sigma_{t|t-1} &\approx G_{t|t-1}^2\omega_{t|t-1} + \sigma^2
\end{aligned}$$

Refined approximations of $g\left(x_t\right)$ may perform better, but require the calculation of additional, higher-order derivatives.

[U] Updating predictions works exactly the same as for the KF, except the Kalman gain is approximated. Note that

$$\mathbf{cov}\left(g(x_t), x_t|\mathcal{F}_{t-1}\right) = G_{t|t-1}\omega_{t|t-1} + \mathbf{cov}(r_t, x_t|\mathcal{F}_{t-1})$$

so that if $r_t \approx 0$, the Kalman gain may be approximated using

$$K_{t|t-1} \approx G_{t|t-1}\omega_{t|t-1}(G_{t|t-1}^2\omega_{t|t-1} + \sigma^2)^{-1}.$$

In addition to the approximation bias, linear updating is no longer optimal in the sense that linear projections do not optimally estimate $p(x_t|y_t, \mathcal{F}_{t-1})$ when the model is nonlinear. This non-optimality is a second potential source of bias. In particular, the functional form of $G(x)$ is integrable, $G_{t|t-1}$ goes to zero as $x_{t|t-1}$ goes to ∞ or $-\infty$. In other words, the information of new observation, y_t is ignored. This pitfalls of EKF become more severe when latent variable follows I(1) process.

3. Unscented Kalman Filter

A linear transformation of a normal distribution may be identified by the mean and variance of the untransformed and transformed (normal) distributions. For a linear approximation of a nonlinear transformation, the mean and variance of the transformed distribution may be meaningless, unless the approximation is good – i.e., if the nonlinear function is "close" to linear. The idea of the UKF is to parsimoniously and nonrandomly select points from the untransformed distribution, pass them through the nonlinear function, and then use the transformed points to make inferences about the functional parameters.

Estimation similarly proceeds in three steps. Our discussion is specific to the model given above, with a univariate, autoregressive state. For a more general and more detailed discussion of the UKF, the reader is referred to Julier and Uhlmann (1997), Julier, Uhlmann, and Durrant-Whyte (1993), or Simon (2006).

[P] A set of three sigma points $\xi_{t-1,i}$ for $i = 0, 1, 2$ are chosen, given by

$$\begin{aligned}
\xi_{t-1,0} &= x_{t-1|t-1} \\
\xi_{t-1,1} &= x_{t-1|t-1} + \sqrt{(1+\lambda)\,\omega_{t-1|t-1}} \\
\xi_{t-1,2} &= x_{t-1|t-1} - \sqrt{(1+\lambda)\,\omega_{t-1|t-1}}
\end{aligned}$$

where λ is a tuning parameter. Using

$$\xi_{t,i} = \xi_{t-1,i},$$

the prediction equations are given by

$$\begin{aligned}
x_{t|t-1} &= \sum\nolimits_{i=0}^{2} w_i \xi_{t,i} \\
\omega_{t|t-1} &= \sum\nolimits_{i=0}^{2} w_i \left(\xi_{t,i} - x_{t|t-1}\right)^2 + 1
\end{aligned}$$

where

$$w_0 = \lambda/(1+\lambda) \qquad \text{and} \qquad w_i = 1/(2(1+\lambda)) \qquad \text{for } i = 1, 2$$

are weights given to the sigma points.

[L] These points are then propagated through the nonlinear function g, so that

$$\zeta_{t,i} = g(\xi_{t,i})$$

and

$$\begin{aligned} y_{t|t-1} &= \sum_{i=0}^{2} w_i \zeta_{t,i} \\ \sigma_{t|t-1} &= \sum_{i=0}^{2} w_i (\zeta_{t,i} - y_{t|t-1})^2 + \sigma^2 \end{aligned}$$

are used to calculate the conditional likelihood.

[U] Updating is still linear and proceeds as above using (4.4), with

$$K_{t|t-1} = \sigma_{t|t-1}^{-1} \sum_{i=0}^{2} w_i (\xi_{t,i} - x_{t|t-1})(\zeta_{t,i} - y_{t|t-1})$$

as the Kalman gain. Note that, like the EKF, linear updating of the UKF is not optimal.

4. Density-Based Nonlinear Filter

The DNF is broadly used for filters that uses the relationship of (4.1),(4.2) and (4.3) instead of relying on approximation of nonlinear function or distribution. Since The DNF is to deal with whole distribution information instead of the first and second moments because I do not assume the normality of the conditional density function. Here I denote more specifically DNF as Numerical Integration Filter with deterministic nodes. In Tanizaki (1996) Numerical Integration Filter uses random draws from

the constructed intervals by the mean and variance of the latent variables based on Extended Kalman filtering.[1] I choose the simple equally-divided nodes for the fixed length of interval. Since the latent variables of unit-root or near unit-root process can go anywhere, however, the location of the interval should adjust depending the updated expected values of the latent variable for each time. The reader is referred to Tanizaki (1996) for more details and an excellent survey of nonlinear filtering techniques.

A disadvantage is that the DNF may slow the implementation of the filter substantially. For this reason, the UKF was developed by engineers concerned with high-speed tracking applications. For econometric applications, the speed of the filter is not usually important.

[P] Using the change in variables, i.e., $x_{t-1} = z + \rho x_{t-2|t-2}$

$$\begin{aligned}
p(x_t|\mathcal{F}_{t-1}) &= \int p(x_t|x_{t-1})p(x_{t-1}|\mathcal{F}_{t-1})\,dx_{t-1} \\
&\approx \int_{-c+\rho x_{t-2|t-2}}^{c+\rho x_{t-2|t-2}} p(x_t|x_{t-1})p(x_{t-1}|\mathcal{F}_{t-1})\,dx_{t-1} \\
&= \int_{-c}^{c} p(x_t|z+\rho x_{t-2|t-2})p(z+\rho x_{t-2|t-2}|\mathcal{F}_{t-1})\,dz \\
&\approx \frac{h}{\sqrt{2\pi}} \sum_{j=1}^{m} \exp\left(-\frac{1}{2}\Big[x_t - \rho(z_j + \rho x_{t-2|t-2})\Big]^2\right) p(z_j + \rho x_{t-2|t-2}|\mathcal{F}_{t-1})
\end{aligned}$$

where c is the one-side length of interval for the conditional density function to be measured, h is the length of each partition in the interval, $j = 1, 2, \cdots, m$, $m = 2c/h$ so that $z = [-c, -c+h, \cdots, c-h]$. Finally the relation $x_t =$

[1] But what if Extended Kalman filtering works bad? In our model Extended Kalman filtering has bigger possibility of being stuck with higher or lower values of latent variable due to the integrability of logistic function. Once it become stuck, so is the intervals.

$z + \rho x_{t-1|t-1}$ yields

$$p(z_i + \rho x_{t-1|t-1}|\mathcal{F}_{t-1}) \approx \frac{h}{\sqrt{2\pi}} \sum_{j=1}^{m} \exp\left(-\frac{1}{2}\Big[z_i + \rho x_{t-1|t-1} - \rho(z_j + \rho x_{t-1|t-2}) \Big]^2 \right) p(z_j + \rho x_{t-2|t-2}|\mathcal{F}_{t-1})$$

Note that $p(z + \rho x_{t-2|t-2}|\mathcal{F}_{t-1})$ is the density shifted from $p(x_{t-1}|\mathcal{F}_{t-1})$ by $\rho x_{t-2|t-2}$.

[L] Using the change in variables, i.e., $x_t = z + \rho x_{t-1|t-1}$

$$\begin{aligned} p(y_t|\mathcal{F}_{t-1}) &= \int p(y_t|x_t) p(x_t|\mathcal{F}_{t-1})\, dx_t \\ &\approx \int_{-c+\rho x_{t-1|t-1}}^{c+\rho x_{t-1|t-1}} p(y_t|x_t) p(x_t|\mathcal{F}_{t-1})\, dx_t \\ &= \int_{-c}^{c} p(y_t|z + \rho x_{t-1|t-1}) p(z + \rho x_{t-1|t-1}|\mathcal{F}_{t-1})\, dz \\ &\approx \frac{h}{\sqrt{2\pi}} \sum_{j=1}^{m} \exp\left(-\frac{1}{2}\Big[y_t - f(z_j + \rho x_{t-1|t-1}) \Big]^2 \right) p(z_j + \rho x_{t-1|t-1}|\mathcal{F}_{t-1}) \end{aligned}$$

Note that $p(z+\rho x_{t-1|t-1}|\mathcal{F}_{t-1})$ is the density shifted from $p(x_t|\mathcal{F}_{t-1})$ by $\rho x_{t-1|t-1}$.

[U] Using the change in variables, i.e., $x_t = z + \rho x_{t-1|t-1}$

$$p(z_i + \rho x_{t|t-1}|\mathcal{F}_t) \approx \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\Big[y_t - f(z_i + \rho x_{t-1|t-1}) \Big]^2 \right) \frac{p(z_i + \rho x_{t-1|t-1}|\mathcal{F}_{t-1})}{p(y_t|\mathcal{F}_{t-1})}$$

C. Experimental Design

1. Simulation

As a framework for our Monte Carlo simulations, I construct $R = 1,000$ pseudo-randomly generated samples of length $n = 1,000$ four each of three different nonlinear functions, with either $\rho = 0.5$ or $\rho = 1.0$. The different values of ρ correspond to and are representative of I(0) and I(1) autoregressive states. Examination of both

Table 14.: Logistic Function Parameters.

Case	ρ	ν	α	σ
ES05	0.5	0.5	2.0	1.0
EN05	1.0	0.5	2.0	1.0
ES50	0.5	5.0	2.0	1.0
EN50	1.0	5.0	2.0	1.0

of these common types of autoregressive processes is important, since the underlying time series characteristics of the state may significantly affect both the time series characteristics of the observed series and also of the parameter estimators.

Similarly to the target zone exchange rate model of Miller and Park (2008), I first consider a family of logistic functions parameterized by

$$g\left(x_t, \theta\right) = \nu / \left(1 + \exp\left(-\alpha x_t\right)\right)$$

with $\nu, \alpha > 0$. Note that the unit variance restriction on (v_t) is critical for identifying α.

Identification of α may still be difficult – especially with integrated (x_t). As Granger and Terasvirta (1993) noted, clustering of observations near the bounds of the logistic function should make the upper bound ν easy to identify, but the slope α difficult. They suggest that, under empirical uncertainty about this parameter, one should conclude that it is in fact large, suggesting a steep slope or sudden transition between upper and lower bound. In order to try to pinpoint such an identification problem, I consider both a relatively small value of ν (where identification of α may be difficult), and relatively large value of ν (where identification of α may be less difficult). Specifically, for $(\rho, \theta, \sigma^2) = (\rho, \nu, \alpha, \sigma^2)$, I consider four cases, which I will refer to as cases ES05, EN05, ES50, EN50. The true parameter values for these cases are given in Table 14.

Table 15.: Power Function Parameters.

Case	ρ	β	γ	σ
PS05	0.5	1.0	0.5	0.02
PN05	1.0	1.0	0.5	0.02

I also consider a power function parameterized by

$$g(x_t, \theta) = \beta x_t^{\gamma}$$

with $\gamma > 0$. Two cases considered for $(\rho, \theta, \sigma^2) = (\rho, \beta, \gamma, \sigma^2)$ are summarized in Table 15.

I consider smaller $\sigma = 0.02$ simply to increase the signal-to-noise ratio in the simulated models.

Since our true parameter value of γ is 0.5, our x_t should be positive for all t. So I can use absolute or exponential or logistic transformation for x_t. Here I use logistic transformation because absolute transformation is not one-to-one mapping and exponential transformation has identification problem between γ and coefficient inside exponential function. However when I use logistic transformation I reduce the nonstationarity of transformed latent variable when original latent variable is nonstationary.

2. Estimation

For simulations, I estimate ρ with the restriction that $\rho \in (-1, 1)$. Even when $\rho = 1$, I can still get the estimation results close enough to 1 due to logistic restriction. I restrict $x_0 = 0$ in both simulation and estimation. I restrict the variance of (v_t) to be unity in both simulation and estimation. As a result of experimentation, I set $\omega_0 = 0.1$ for the EKF and UKF and, for the UKF, I set the tuning parameter $\lambda = 1/3$

and, for the DNF, I set the tuning parameter c = 5, h = 0.1. I do not smooth any of the estimated series $\left(x_{t|t}\right)$.

3. Criteria for Comparison

I compare the four filters across each of these six cases in four primary dimensions: parameter estimation, state estimation, in-sample fit of the model, and numerical stability.

- **Parameter estimation.** An obvious metric for evaluating accuracy of these techniques is to examine parameter estimates and their standard errors. I can only compare the EKF, UKF, and DNF using this criterion, since the KF assumes an incorrect (linear) functional form.

- **State estimation.** I analyze the fit of the estimated state $\left(x_{t|t}\right)$ by examining the distribution across simulations of

$$RMSE_r\left(x_{t|t}\right) = \sqrt{n^{-1}\textstyle\sum_t \left(x_{t|t} - x_t\right)^2},$$

  the root MSE across $t = 1, \ldots, n$ for each simulation $r = 1, \ldots, R$.

- **In-sample fit of the model.** I analyze the in-sample fit of the model by examining the distribution across simulations of

$$RMSE_r\left(y_{t|t-1}\right) = \sqrt{n^{-1}\textstyle\sum_t \left(y_{t|t-1} - y_t\right)^2},$$

  the model root MSE for each simulation r. Since (x_t) is not observable, comparisons based on the in-sample fit of the model may be quite different from comparisons based on the parameter estimates themselves. For example, if parameter estimates are particularly bad, it is possible that $\left(x_{t|t}\right)$ is still constructed in such a way that $(g(x_{t|t}, \hat{\theta}))$ is still a good fit for (y_t).

- **Numerical stability of the algorithm.** Numerical stability is generally problematic in estimation of nonlinear models, and is particularly problematic in nonlinear filtering applications. Nonlinearity may create extreme values in simulation, identification problems, and other numerical eccentricities that may cause the algorithm to fail. In our experiments, I discard simulations in which the algorithm fails. I use the same seed for all four techniques until I get more than $1,000$ successful simulations for all techniques. I discard unsuccessful simulations. Then, I *randomly* discard successful simulations until I have exactly $1,000$ for each technique. However, the failures themselves convey information about the numerical stability of the respective techniques. I therefore report the percentage of successful simulations out of total simulations.

D. Experimental Results and Conclusions

Table 16 and Table 17 summarize the main results of the analysis. The odd panels show the results for simulated I(1) states, while the even panels show those for I(0) states. Within each panel, I first show the mean and standard deviation of parameter estimates across all simulations. These standard deviations are not standard errors of a single estimate, but they may be interpreted in a roughly similar manner. Specifically, they provide a degree of certainty (or uncertainty) about the mean of the parameter estimates. Since I know the true values used for our simulations, I may assess how well the estimates target these values. Below them, I display the mean and median of $RMSE_r\left(x_{t|t}\right)$ and $RMSE_r\left(y_{t|t-1}\right)$, providing a measure of the accuracy of state estimation and of in-sample fit of the model, respectively. Finally, stability is assessed with the percentage of the total simulations in which numerical optimization failed.

Table 16.: Numerical Results with Logistic Function.

EN05	DNF		UKF		EKF		KF	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
$\rho = 1.0$	0.4656	0.3820	0.8216	0.3236	0.8456	0.3131	NA	NA
$\nu = 0.5$	0.6695	0.3228	0.5949	0.2976	0.5302	0.3125	NA	NA
$\alpha = 2.0$	9.9128	13.4138	5.0362	9.2182	1.4373	2.0734	NA	NA
$\sigma = 1.0$	0.9803	0.0352	0.9932	0.0308	0.9669	0.1563	NA	NA
	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.
RMSE(x)	20.0040	17.5399	17.2164	13.7640	17.8033	14.4902	20.6737	16.3798
RMSE(y)	1.0109	1.0105	1.0082	1.0081	1.0074	1.0080	1.0081	1.0082
Stability	52.1%		57.4%		82.6%		84.0%	

ES05	DNF		UKF		EKF		KF	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
$\rho = 0.5$	0.2262	0.4358	0.5407	0.3924	0.4711	0.3880	NA	NA
$\nu = 0.5$	0.5061	0.0652	0.5038	0.0675	0.4996	0.0685	NA	NA
$\alpha = 2.0$	13.8948	14.8334	7.0522	15.5002	1.9422	2.2186	NA	NA
$\sigma = 1.0$	0.9907	0.0251	1.0007	0.0254	0.9188	0.2180	NA	NA
	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.
RMSE(x)	1.1429	1.1409	1.3614	1.1629	1.3139	1.1773	4.2359	1.1673
RMSE(y)	1.0134	1.0131	1.0135	1.0138	1.0064	1.0073	1.0128	1.0126
Stability	60.8%		53.3%		75.4%		72.3%	

EN50	DNF		UKF		EKF		KF	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
$\rho = 1.0$	0.9854	0.0547	0.9570	0.1012	0.9845	0.0240	NA	NA
$\nu = 5.0$	4.9648	0.4326	5.5545	1.9609	4.9253	1.3327	NA	NA
$\alpha = 2.0$	2.8510	5.2895	1.6676	2.7398	1.0202	1.5647	NA	NA
$\sigma = 1.0$	0.9928	0.0245	1.0373	0.0490	1.0339	0.0377	NA	NA
	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.
RMSE(x)	13.1901	10.3225	13.7505	9.7338	14.6859	11.2501	19.2012	15.5687
RMSE(y)	1.1318	1.1219	1.1538	1.1494	1.1143	1.1134	1.1768	1.1738
Stability	92.8%		66.8%		89.2%		93.3%	

ES50	DNF		UKF		EKF		KF	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
$\rho = 0.5$	0.4994	0.0414	0.5635	0.0582	0.5084	0.0492	NA	NA
$\nu = 5.0$	4.9973	0.1330	5.0539	0.1738	5.0196	0.1208	NA	NA
$\alpha = 2.0$	2.0142	0.1615	2.6226	0.7394	1.1446	0.0954	NA	NA
$\sigma = 1.0$	0.9983	0.0484	1.1056	0.1994	1.1463	0.1068	NA	NA
	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.
RMSE(x)	0.6338	0.6328	0.7550	0.7528	0.6445	0.6431	3.7616	1.2893
RMSE(y)	1.8237	1.8243	1.8270	1.8282	1.3054	1.3044	1.8245	1.8252
Stability	100.0%		99.9%		100.0%		100.0%	

Table 17.: Numerical Results with Power Function.

PN05	DNF		UKF		EKF		KF	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
$\rho = 1.0$	0.9964	0.0092	0.9984	0.0034	0.9959	0.0072	NA	NA
$\beta = 1.0$	0.9995	0.0406	1.0061	0.0740	1.0677	0.1562	NA	NA
$\gamma = 0.5$	0.5181	0.0720	0.5298	0.0992	0.6163	0.2225	NA	NA
$\sigma = 0.02$	0.0198	0.0018	0.0194	0.0034	0.0158	0.0083	NA	NA
	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.
RMSE(x)	2.6136	1.9047	2.9852	2.3116	4.6102	2.5127	26.2282	20.3418
RMSE(y)	0.0277	0.0283	0.0270	0.0271	0.0284	0.0287	0.0384	0.0384
Stability	33.7%		63.4%		26.0%		94.4%	

PS05	DNF		UKF		EKF		KF	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
$\rho = 0.5$	0.4962	0.0911	0.4734	0.1054	0.4970	0.0982	NA	NA
$\beta = 1.0$	1.0072	0.0628	1.7710	0.8573	1.0074	0.0676	NA	NA
$\gamma = 0.5$	0.5077	0.0889	1.2212	0.5012	0.5080	0.0952	NA	NA
$\sigma = 0.02$	0.0193	0.0032	0.0185	0.0040	0.0191	0.0039	NA	NA
	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.
RMSE(x)	0.7866	0.7866	0.9001	0.9001	0.7895	0.7895	1.5553	1.5553
RMSE(y)	0.0275	0.0275	0.0275	0.0275	0.0275	0.0275	0.3666	0.3747
Stability	73.0%		56.2%		94.2%		98.8%	

Obviously, since the KF is inherently misspecified, it cannot estimate the parameters of the nonlinear models. Among the remaining three techniques, it is somewhat surprising to learn that there is no clear leader. As discussed above, the EKF has two sources of bias. The UKF eliminates one source, and the DNF eliminates both. I may therefore expect that the DNF dominates the UKF, which dominates the EKF. While the EKF is usually dominated by one or both of the other two, it is not consistently dominated by one of them.

The results for parameter estimates in fact depend on the functions and parameters estimated. Consider first the logistic specifications with a relatively small ceiling (EN05 and ES05). I can expect from the observation made by Granger and Terasvirta (1993) that the slope α will be very difficult to identify. This certainly appears to be the case, since the mean of the estimates of α are far off (except for the EKF), and the standard deviations of the estimates are very large for all three techniques. The uncertainty about α causes much uncertainty about ρ, with the DNF performing the worst – in the unit root case, $\rho = 1$ is not even within a standard deviation of the mean. However, all of the techniques estimate ν and σ reasonably well.

In order to avoid the identification problem (at least in these finite samples), I also examine cases EN50 and ES50, with much larger ceilings. The slope α may be traced out more effectively, since more observations will rest between the floor and the ceiling of the function. Of course, this should hold more so for ES50 than EN50, because of the strong mean reversion of an I(0) process. In the I(1) case, there is less uncertainty than with the lower ceiling, but still evident identification problems. The performance of the parameter estimation would be well compared if one see the densities of the estimates. Figure 18 and Figure 19 indicates that in the any case DNF outperforms the other filters. And in the I(1) case, EKF does not work worse because there is the possibility that the Kalman gain disappears when the latent variables

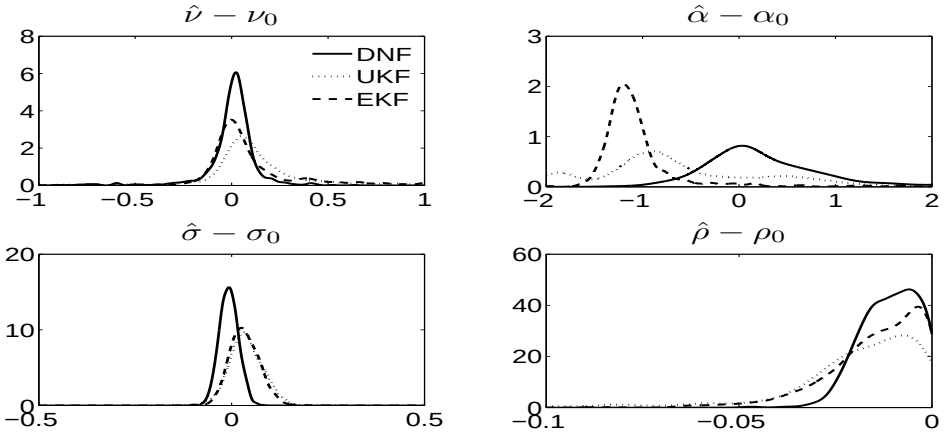

Fig. 18.: Densities of the EN50 Estimates with Four Filters.

stay in the extreme region.

Looking at a completely different functional form, the power function (PN05 and PS05), the DNF slightly outperforms the other two for both I(1) and I(0) cases. This fact is also confirmed in the Figure 20 and Figure 21. All of them appear to perform reasonable well, except for the UKF in the stationary case. In particular, the UKF does not estimate the power γ adequately.

Turning to state estimation, I can now consider the KF with the other three. However, it is no surprise that estimating a state that enters into the model nonlinearly with a technique that assumes linearity will generally not work well. Indeed, this seems to be borne out in our results. There is an interesting exception, however, which is EN05. Recall that this is the case in which α is the least identifiable. In large samples, this model is approximately piecewise linear (with no slope). It is therefore not surprising that the KF is competitive with the other techniques at estimating the state. In fact, judging from a similar mean a lower median of the RMSE, the KF actually does a *better* job than the DNF. In both cases with logistic function and I(1) state, the UKF estimates the state most accurately, with lower mean and median

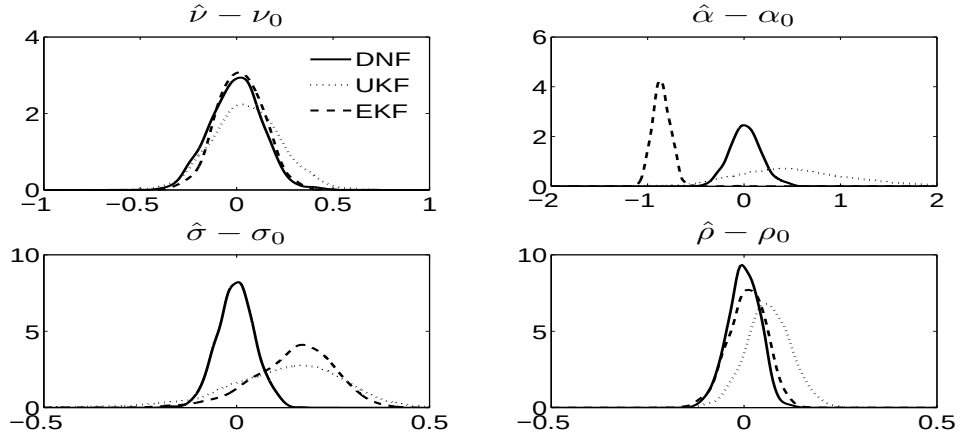

Fig. 19.: Densities of the ES50 Estimates with Four Filters.

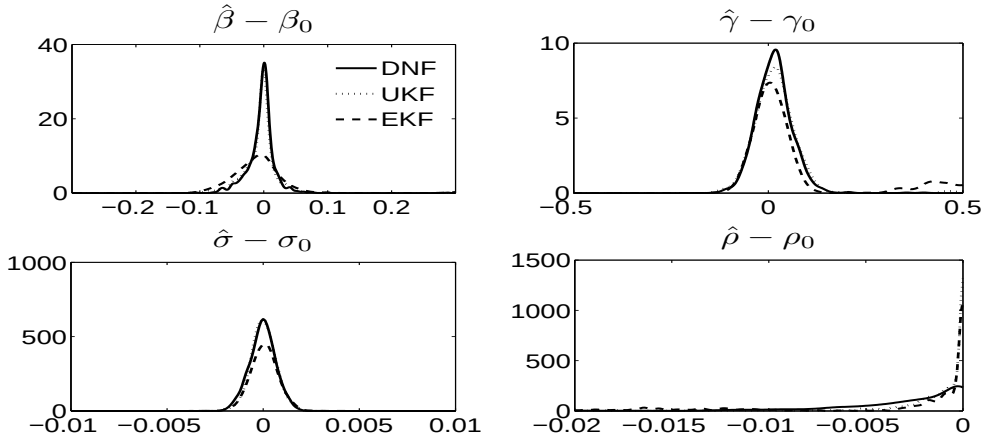

Fig. 20.: Densities of the PN50 Estimates with Four Filters.

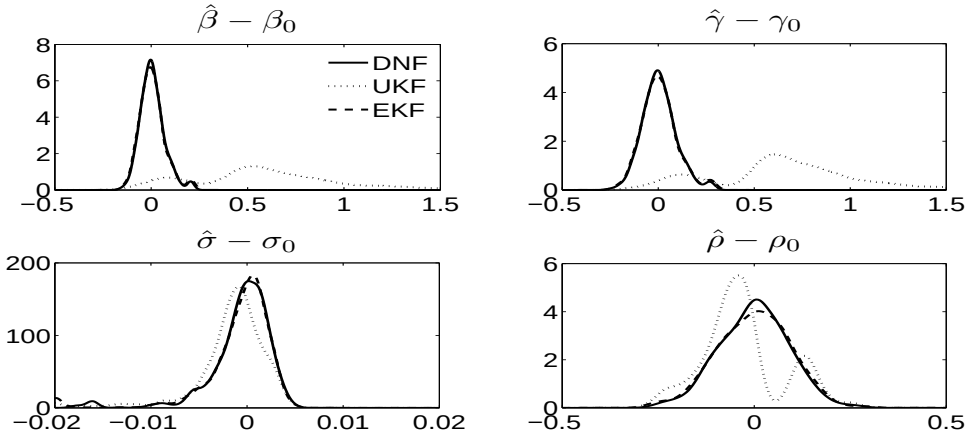

Fig. 21.: Densities of the PS50 Estimates with Four Filters.

RMSE. However, the DNF outperforms the UKF in all of the other cases. In some of the stationary cases, the EKF also dominates the UKF.

All four techniques – even the KF – are competitive at overall in-sample fit. The EKF turns out to be the best at fitting logistic functions, but all three of the nonlinear techniques fit the power function nearly equally well. The KF does not lag far behind, and – again, in the exceptional asymptotically piecewise linear case – outperforms the DNF in this dimension.

Finally, the numerical stability of the KF is clearly unrivaled, as it does not have to contend with any of numerical problems that can inherently occur with nonlinear techniques. With one exception, the EKF is almost as stable. The UKF and DNF appear to be less stable than the KF and EKF in most cases.

Overall, the results of our numerical experiments do not favor any of the filters over the others in every case. If parameter estimates of a well-identified nonlinear model are needed, then the DNF appears preferable. However, if the parameters are not well-identified, the DNF appears inferior. If the goal is to filter the state, then the UKF appears to the best for an I(1) state, although the DNF appears to be the best

for an I(0) state. If in-sample fit of the model is the overarching goal of estimating a nonlinear state space model, it appears that the EKF is the best choice. Finally, for numerical stability, the KF and EKF are best in most cases.

CHAPTER V

CONCLUSION

We construct nonlinear state-space model with persistent latent variables to investigate the volatility generating process from the return or growth data of an asset. We argue that a logistic function has some advantages over standard exponential function for the non-explosive volatility and asymmetric leverage effect. Due to the nonlinearity of logistic volatility function and persistence of the latent variables, conventional Extended Kalman filtering is not appropriate with our model. To filter the volatility generating process from data we introduce two alternative methods, which are density-based ML estimation and Gibbs sampling. These two methods give similar results though only density-based ML estimation allows extensive Monte-Carlo Experiments. We find that all the parameter can be correctly estimated from Monte-Carlo Experiments. We apply our methods to stock return and dividend growth data and extract the volatility generating process. The extracted volatility generating process explain the realized volatility quite well and moreover we can find some fundamentals cointegrated with them.

We extend our model into multivariate setup for extracting common stochastic trend, in particular, we introduce a stochastic volatility model with consumption and dividend process to identify macroeconomic uncertainty. With this extended model, we try to link the macroeconomic uncertainty and asset pricing because macroeconomic uncertainty, which is time-varying but unobservable, is considered as an important ingredient for asset pricing. We solve this model numerically with Bayesian approach to avoid the multidimensional difficulties. We find that the extracted volatil-

ity series explains well the realized volatility series of both consumption and dividend. Also we see a counter-cyclical relation of the extracted macroeconomic uncertainty.

And then motivated from Bansal and Yaron (2004), we combine this stochastic volatility model with long-run risk model and Epstein-Zin-Weil preference. We find our estimated risk-aversion coefficient and intertemporal elasticity of substitution around two which is plausible according to the consensus in the finance literature. Bansal and Yaron's model with relatively high risk aversion can generate high risk premium through the persistent long-run risk channel. However, our model produces high risk premium even with moderate coefficient of risk aversion because it has another channel, more realistic time-varying volatility. Furthermore, we find that the market return is as volatile as real data and the risk-free rate is low and stable.

REFERENCES

An, X. (2007): "Macroeconomic Conditions, Systematic Risk Factors, and the Time Series Dynamics of Commercial Mortgage Credit Risk," SSRN Working Paper Series, http://ssrn.com/abstract=972614.

Attanasio, P. O., and G. Weber (1989): "Intertemporal Substitution, Risk Aversion and the Euler Equation for Consumption," *Economic Journal*, 99, 59–73.

Bacchetta, P., and S. Gerlach (1997): "Consumption and Credit Constraints: International Evidence," *Journal of Monetary Economics*, 40, 207–238.

Bansal, R., D. Kiku, and A. Yaron (2007): "Risks for the Long Run: Estimation and Inference," Working Paper, Duke University.

Bansal, R., and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59(4), 1481–1509.

Barro, R. J. (2005): "Rare Disasters and Aseet Markets in the Twentieth Century," *Quarterly Journal of Economics*, 121(3), 823–866.

Bijleveld, F., J. Commandeur, S. J. Koopman, and K. van Montfort (2007): "Multivariate Nonlinear Time Series Modelling of Exposure and Risk in Road Safety Research," Working Paper, Vrije Universiteit, Mimeo.

Black, F. (1976): "Studies of Stock Market Volatility Change," *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pp. 177–181.

Bolland, P. J., and J. T. Connor (1997): "A Constrained Neural Network Kalman Filter for Price Estimation in High Frequency Financial Data," *International Journal of Neural System*, 8, 399–415.

Campbell, J. Y. (2002): "Consumption-Based Asset Pricing," Harvard Institute Research Working Paper No.1974.

Campbell, J. Y., and J. H. Cochrane (1999): "By Force of Habit: A Consumption-based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107, 205–251.

Carter, C. K., and R. Kohn (1994): "On Gibbs Sampling for State Space Models," *Biometrika*, 81(3), 541–553.

Chang, Y., J. I. Miller, and J. Y. Park (2006): "Extracting a Common Stochastic Trend: Theory with Some Application," *Journal of Econometrics*, forthcoming.

Chen, R.-R., and L. Scott (2003): "Multi-factor Cox-Ingersoll-Ross Models of the Term Structure: Estimates and Tests from a Kalman Filter Model," *Journal of Real Estate Finance and Economics*, 27, 143–172.

Chib, S., and E. Greenberg (1995): "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49(4), 327–335.

Constantinides, G. M., J. B. Donalson, and R. Mehra (2002): "Junior Can't Borrow: A New Perspective On the Equity Premium," *Quarterly Journal of Economics*, 117, 269–296.

Duan, J.-C., and J.-G. Simonato (1997): "Estimating and Testing Exponential-

Affine Term Structure Models by Kalman Filter," *Review of Quantitative Finance and Accounting*, 13, 111–135.

Duffee, G. R. (1999): "Estimating the Price of Default Risk," *Review of Financial Studies*, 12, 197–226.

Epstein, L., and S. Zin (1989): "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Aseet Return: A Theoretical Framework," *Econometrica*, 57, 937–969.

Geman, S., and D. Geman (1984): "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern and Machine Intelligence*, 6(6), 721–741.

Geweke, J. (1992): "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in *Bayesian Statistics*, ed. by J. Berger, J. Bernardo, A. Dawid, and A. Smith. Vol. 4, pp. 169-194, Oxford: Oxford University Press.

Geweke, J., and H. Tanizaki (2001): "Bayesian Estimation of State-Space Models Using the Metropolis-Hastings Algorithm within Gibbs Sampling," *Computational Statistics and Data Analysis*, 37, 151–170.

Geyer, A. L., and S. Pichler (1999): "A State-Space Approach to Estimate and Test Multi-Factor Cox-Ingersoll-Ross Models of the Term Structure," *Journal of Financial Research*, 22, 107–130.

Granger, C. W., and T. Terasvirta (1993): *Modeling Nonlinear Economic Relationships*. Oxford: Oxford University Press.

Grillenzoni, C. (1993): "ARIMA Processes with ARIMA Parameters," *Journal of Business and Economic Statistics*, 11, 235–250.

Hamilton, J. D., and G. Lin (1996): "Stock Market Volatility and the Business Cycle," *Journal of Applied Econometrics*, 11, 573–593.

Hansen, L. P., and K. Singleton (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

Harvey, A. C. (1990): *Forecasting, Strucutural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, A. C., and N. Shepard (1996): "Estimation of an Asymmetric Stochastic Volatility Models for Asset Returns," *Journal of Business and Economic Statistics*, 14(4), 429–434.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994): "Bayesian Analysis of Stochastic Volatiltiy Model," *Journal of Business and Economic Statistics*, 12(4), 371–389.

——— (2004): "Bayesian Analysis of Stochastic Volatiltiy Models with Fat-Tails and Correlated Errors," *Journal of Econometrics*, 122, 185–212.

Jazwinski, A. H. (1970): *Stochastic Processes and Filtering Theory*. New York: Academic Press.

Julier, S. J., and J. K. Uhlmann (1997): "A New Extension of the Kalman Filter to Nonlinear Systems," in *The Proceedings of AeroSense: The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls*, pp. 182–193. SPIE.

Julier, S. J., J. K. Uhlmann, and H. F. Durrant-Whyte (1993): "A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators," *IEEE Transactions on Automatic Control*, 45, 477–482.

Kalman, R. E. (1960): "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, 82, 34–45.

Lund, J. (1994): "Econometric Analysis of Continuous-Time Arbitrage-Free Models of the Term Structure of Interest Rates," Working Paper No. 89, Aarhus School of Business.

Mehra, R., and E. C. Prescott (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–161.

Miller, J. I., and J. Y. Park (2008): "Nonlinearity, Nonstationarity, and Thick Tails: How They Interact to Generate Persistency in Memory," Working Paper 08-01, Department of Economics, University of Missouri.

Park, J. Y. (1990): "Testing for Unit Roots and Cointegration by Variable Addition," *Advances in Econometrics*, 8, 107–133.

——— (2002): "Nonstationary Nonlinear Heteroskedasticity," *Journal of Econometrics*, 110, 383–415.

Rietz, T. A. (1988): "The Equity Risk Premium: A Solution," *Journal of Monetary Economics*, 22, 117–131.

Shen, C.-H., D. R. Hakes, and K. Brown (1999): "Time-Varying Response of Monetary Policy to Macroeconomic Conditions," *Southern Economic Journal*, 65, 584–593.

Simon, D. (2006): *Optimal State Estimation: Kalman, H_∞ and Nonlinear Approaches*. Hoboken, NJ: John Wiley & Sons.

So, M. K. P., K. Lam, and W. K. Li (1998): "A Stochastic Volatility Model with Markov Switching," *Journal of Business and Economic Statistics*, 16(2), 244–253.

Tanizaki, H. (1993): "Kalman Filter Models with Qualitative Dependent Variables," *The Review of Economics and Statistics*, 75, 747–752.

Tanizaki, H. (1996): *Nonlinear Filters: Estimation and Application*. 2nd, New York: Springer.

Tanner, M. A., and W. H. Wong (1987): "The Calculation of Posterior Distributions By Data Augmentation," *Journal of the American Statistical Association*, 82, 528–549.

Tierney, L. (1994): "Markov Chain for Exploring Posterior Distributions," *Annals of Statistics*, 22(4), 1701–1762.

Vissing-Jorgensen, A. (2002): "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution," *Journal of Political Economy*, 110, 825–853.

Weil, P. (1989): "The Equity Premium Puzzle and the Risk-free Rate Puzzle," *Journal of Monetary Economics*, 24, 401–421.

Yu, J. (2005): "On the Leverage in a Stochastic Volatility Model," *Journal of Econometrics*, 127, 165–178.

APPENDIX A

SOLUTION TO THE RETURNS ON THE ALL INVESTED WEALTH

We start from the Euler equation for the all invested wealth $R_{w,t+1}$.

$$E_t\left[\exp(\chi \ln\delta - \frac{\chi}{\psi}g_{c,t+1} + \chi r_{w,t+1})\right] = 1$$

Plug in the consumption growth process $g_{c,t+1}$, and approximate $r_{w,t+1} \approx k_{0,c} + k_{1,c}z_{c,t+1} - z_{c,t} + g_{c,t+1}$. And we conjecture the log price-consumption ratio $z_{c,t} \equiv \log\left(\frac{P_{c,t}}{C_t}\right)$ is dependent on the state variable $\eta_{c,t}$ and $f_{c,t}$, such that $z_{c,t} = A_{0,c} + A_{1,c}\eta_{c,t} + A_{2,c}f_{c,t}$. We use $f_{c,t}$ to denote $f_c(\lambda_c w_t)$ for short-hand notation.

$$E_t\left[\exp\left(\chi \ln\delta + (\chi - \frac{\chi}{\psi})g_{c,t+1} + \chi\{k_{0,c} + k_{1,c}z_{c,t+1} - z_{c,t}\}\right)\right] = 1$$

And we ignore the difference between $f_c(x_{c,t})$ and $f_c(\lambda_c w_t)$ because $x_{c,t}$ and w_t are cointegrated each other and in particular, logistic transformation makes two values similar in the high or low volatility regime at which mass will cluster asymptotically. Moreover consumption volatility moves very small range between 0.000002 and 0.000022. Following Bansal and Yaron(2004), we will verify the unknown coefficients for $z_{c,t}$ process.

First, we can linearize $f_{c,t+1}$ around $f_{c,t} \approx f_{c,0}$.

$$\begin{aligned}
f_{c,t+1} &= \alpha_c + \frac{\beta_c\left(f_{c,t+1} - \alpha_c\right)}{f_{c,t+1} - \alpha_c + (\alpha_c + \beta_c - f_{c,t})\exp(-\lambda_c u_{t+1})} \\
&\approx \alpha_c + \frac{\beta_c\left(f_{c,0} - \alpha_c\right)}{f_{c,0} - \alpha_c + (\alpha_c + \beta_c - f_{c,0})\exp(-\lambda_c u_{t+1})} \\
&+ \frac{\beta_c^2\exp\left(-\lambda_c u_{t+1}\right)}{\left[f_{c,0} - \alpha_c + (\alpha_c + \beta_c - f_{c,0})\exp(-\lambda_c u_{t+1})\right]^2}\left[f_{c,t} - f_{c,0}\right] \\
&= \alpha_c + K_1\left(u_{t+1}\right) + K_2\left(u_{t+1}\right)\left[f_c\left(\lambda_c w_t\right) - f_{c,0}\right]
\end{aligned}$$

with $K_1(u_{t+1})$ linearized again around $u_{t+1} \approx 0$

$$\begin{aligned} K_1(u_{t+1}) &\approx \frac{\beta_c(f_{c,0}-\alpha_c)}{f_{c,0}-\alpha_c+(\alpha_c+\beta_c-f_{c,0})} + \frac{\lambda_c\beta_c(f_{c,0}-\alpha_c)(\alpha_c+\beta_c-f_{c,0})}{[f_{c,0}-\alpha_c+(\alpha_c+\beta_c-f_{c,0})]^2}u_{t+1} \\ &= f_{c,0}-\alpha_c+\frac{\lambda_c}{\beta_c}(f_{c,0}-\alpha_c)(\alpha_c+\beta_c-f_{c,0})u_{t+1} \end{aligned}$$

and with $K_2(u_{t+1})$ linearized around $u_{t+1} \approx 0$

$$\begin{aligned} K_2(u_{t+1}) &\approx \frac{\beta_c^2}{[f_{c,0}-\alpha_c+(\alpha_c+\beta_c-f_{c,0})]^2} + \frac{-\lambda_c\beta_c^4+2\lambda_c\beta_c^3(\alpha_c+\beta_c-f_{c,0})}{[f_{c,0}-\alpha_c+(\alpha_c+\beta_c-f_{c,0})]^4}u_{t+1} \\ &= 1+\frac{\lambda_c}{\beta_c}(\beta_c-2f_{c,0}+2\alpha_c)u_{t+1} \end{aligned}$$

Finally, $f_{c,t+1}$ can be expressed in terms of $f_{c,t}$ and u_{t+1}.

$$\begin{aligned} f_{c,t+1} &\approx f_{c,t}+\frac{\lambda_c}{\beta_c}(\beta_c-2f_{c,0}+2\alpha_c)f_{c,t}u_{t+1} \\ &\quad+\frac{\lambda_c}{\beta_c}[(f_{c,0}-\alpha_c)(\alpha_c+\beta_c+f_{c,0})-\beta_c f_{c,0}]u_{t+1} \\ &= f_{c,t}+\Gamma_1 f_{c,t}u_{t+1}+\Gamma_2 u_{t+1} \end{aligned} \tag{A.1}$$

where $\Gamma_1 = \frac{\lambda_c}{\beta_c}(\beta_c-2f_{c,0}+2\alpha_c)$ and $\Gamma_2 = \frac{\lambda_c}{\beta_c}[(f_{c,0}-\alpha_c)(\alpha_c+\beta_c+f_{c,0})-\beta_c f_{c,0}]$. And we can linearize $f_{c,t}^2$ around $f_{c,t} \approx f_{c,0}$

$$f_{c,t}^2 \approx f_{c,0}^2+2f_{c,0}(f_{c,t}-f_{c,0}). \tag{A.2}$$

Using (3.2),(3.3), (3.10), (A.1) and (A.2) and rearranging the inside the exponential of Euler equation (3.1) by constant, $\eta_{c,t}$, and $f_{c,t}$ we have

$$\begin{aligned} &\chi\ln\delta+(1-\gamma)\mu_c+\chi(k_{0,c}+(k_{1,c}-1)A_{0,c})+0.5(\chi k_{1,c}A_{2,c})^2\left[\Gamma_2^2-\Gamma_1^2f_{c,0}^2\right] \\ &+[(1-\gamma)+\chi A_{1,c}(k_{1,c}\rho_c-1)]\eta_{c,t} \\ &+\left[\chi^2k_{1,c}^2\Gamma_1[\Gamma_1f_{c,0}+\Gamma_2]A_{2,c}^2+\chi(k_{1,c}-1)A_{2,c}+\frac{1}{2}\left[(1-\gamma)^2+(\chi k_{1,c}A_{1,c}\varphi_c)^2\right]\right]f_{c,t} \end{aligned}$$

Then when the above equation becomes equal to zero always, coefficients $A_{0,c}$, $A_{1,c}$, and $A_{2,c}$ can be verified as following:

$$A_{1,c} = \frac{1-\gamma}{\chi\left(1-k_{1,c}\rho_c\right)} \tag{A.3}$$

$$A_{0,c} = \frac{\chi \ln\delta + (1-\gamma)\mu_c + \chi k_{0,c} + 0.5\left(\chi k_{1,c}A_{2,c}\right)^2\left[\Gamma_2^2 - \Gamma_1^2 f_{c,0}^2\right]}{\chi\left(1-k_{1,c}\right)} \tag{A.4}$$

$$A_{2,c} = \frac{1-k_{1,c} \pm \sqrt{\left(k_{1,c}-1\right)^2 - 2k_{1,c}^2\Upsilon\left[(1-\gamma)^2 + \left(\chi k_{1,c}A_{1,c}\varphi_c\right)^2\right]}}{2\chi k_{1,c}^2\Upsilon} \tag{A.5}$$

where $\Upsilon = \Gamma_1\left(\Gamma_1 f_{c,0} + \Gamma_2\right)$.

And since inter-temporal marginal rate of substitution(IMRS) is, $m_{t+1} \equiv lnM_{t+1} = \chi\ln\delta - \frac{\chi}{\psi}g_{c,t+1} + (\chi-1)r_{w,t+1}$, conditional mean of IMRS is

$$\begin{aligned} E_t\left(m_{t+1}\right) = {} & \chi\ln\delta - \gamma\mu_c + (\chi-1)\left(k_{0,c} + \left(k_{1,c}-1\right)A_{0,c}\right) \\ & + \left[(\chi-1)\left(k_{1,c}\rho_c - 1\right)A_{1,c}\right]\eta_{c,t} + \left[(\chi-1)A_{2,c}\left(k_{1,c}-1\right)\right]f_{c,t}. \end{aligned}$$

Then the innovations in IMRS is

$$m_{t+1} - E_t(m_{t+1}) = \Lambda_{m\epsilon}\sqrt{f_{c,t}}\epsilon_{c,t+1} - \Lambda_{m,e}\sqrt{f_{c,t}}e_{c,t+1} - \Lambda_{m,u}(t)u_{t+1} \tag{A.6}$$

where the prices of risk for each sources of risk are defined appropriately,

$$\begin{aligned} \Lambda_{m\epsilon} &\equiv -\gamma \\ \Lambda_{m,e} &\equiv (1-\chi)k_{1,c}A_{1,c}\varphi_c \\ \Lambda_{m,u}(t) &\equiv (1-\chi)k_{1,c}A_{2,c}\left(\Gamma_1 f_{c,t} + \Gamma_2\right) \end{aligned}$$

APPENDIX B

SOLUTION TO THE MARKET RETURN

When we log linearize log return on the market assets, $R_{m,t+1} = \frac{D_{t+1}+P_{m,t+1}}{P_{m,t+1}}$ in a similar way, we have

$$r_{m,t+1} \approx k_{0,m}(\bar{z}_m) + k_{1,m}(\bar{z}_m) z_{m,t+1} - z_{m,t} + g_{d,t+1} \tag{B.1}$$

where $z_{d,t} = \log(P_{d,t}/D_t)$ is the log price-dividend ratio.

To solve the market return, we assume the process of dividend growth $g_{d,t+1}$ as (3.13) and the process of long run component of dividend $\eta_{d,t+1}$ as similar as consumption

$$\eta_{d,t+1} = \rho_d \eta_{d,t} + \varphi_d \sqrt{f_d(x_{d,t})} e_{d,t+1} \tag{B.2}$$

We can use the method of undetermined coefficient by guessing approximately $z_{m,t} \approx A_{0,m} + A^c_{1,m}\eta_{c,t} + A^d_{1,m}\eta_{d,t} + A_{2,m} f_{c,t}$. Note that the relevant state variable is $\eta_{c,t}, \eta_{d,t}$ and w_t because dividend process also is affected by consumption. First, approximate $f_d(\lambda_d w_t)$ in terms of $f_{c,t}$ at $f_{d,0}$

$$\begin{aligned} f_d(\lambda_d w_t) &\approx \alpha_d + \frac{\beta_d \left(f_{d,0} - \alpha_c\right)^{\lambda} \exp\left(\lambda\kappa_c - \kappa_d\right)}{\left(f_{d,0} - \alpha_c\right)^{\lambda} \exp\left(\lambda\kappa_c - \kappa_d\right) + \left(\alpha_c + \beta_c - f_{d,0}\right)^{\lambda}} \\ &\quad + \frac{\lambda\beta_c\beta_d \left(f_{d,0} - \alpha_c\right)^{\lambda-1} \left(\alpha_c + \beta_c - f_{d,0}\right)^{\lambda-1} \exp\left(\lambda\kappa_c - \kappa_d\right)}{\left[\left(f_{c,0} - \alpha_c\right)^{\lambda} \exp\left(\lambda\kappa_c - \kappa_d\right) + \left(\alpha_c + \beta_c - f_{d,0}\right)^{\lambda}\right]^2} \left(f_{c,t} - f_{d,0}\right) \\ &= \pi_1 + \pi_2 \left(f_{c,t} - f_{d,0}\right) \end{aligned}$$

where $\lambda = \lambda_d/\lambda_c$ and π_1 and π_2 are defined appropriately.

Here we ignore the difference between $f_d(x_{d,t})$ and $f_d(\lambda_d w_t)$ by the similar reasoning in the Appendix C. To derive expressions for the undetermined coefficients $A_{0,m}$,

$A^c_{1,m}$, $A^d_{1,m}$, and $A_{2,m}$, use the Euler equation again for the market return $r_{m,t+1}$:

$$E_t\left[\exp\left(\chi\ln\delta-\frac{\chi}{\psi}g_{c,t+1}+(\chi-1)\,r_{w,t+1}+r_{m,t+1}\right)\right]=1$$

Expanding $r_{w,t+1}$ and $r_{m,t+1}$ processes, we have

$$E_t\left[\exp\left(\begin{array}{c}\chi\ln\delta-\gamma g_{c,t+1}+(\chi-1)\,(k_{0,c}+k_{1,c}z_{c,t+1}-z_{c,t})\\+k_{0,m}+k_{1,m}z_{m,t+1}-z_{m,t}+g_{d,t+1}\end{array}\right)\right]=1$$

Now we collecting terms inside the exponential by state variables $\eta_{c,t}$, $\eta_{d,t}$, $f_{c,t}$ and constants. Note that though there is cross product term $-\gamma\sqrt{f_{c,t}}\sqrt{f_{d,t}}\rho$ due to the correlation of short-run shocks, empirically both ρ and volatility are so small that this term is negligible. As the same as in the Appendix A, we can find the coefficient terms as following:

$$A_{0,m}=\frac{\chi\ln\delta-\gamma\mu_c+\mu_d+(\chi-1)\,(k_{0,c}+k_{1,c}A_{0,c}-A_{0,c})+k_{0,m}+0.5H_1}{1-k_{1m}}$$

$$A^c_{1,m}=\frac{-\frac{1}{\psi}}{1-k_{1,m}\rho_c}$$

$$A^d_{1,m}=\frac{1}{1-k_{1m}\rho_d}$$

$$A_{2,m}=\frac{[(1-\chi)\,k_{1,c}A_{2,c}k_{1m}\Upsilon+1-k_{1,m}]\pm\sqrt{[(1-\chi)\,k_{1,c}A_{2,c}k_{1m}\Upsilon+1-k_{1,m}]^2-4\Upsilon k^2_{1,m}H_2}}{2\Upsilon k^2_{1,m}}$$

where

$$H_1=\left(1+(k_{1,m}A^d_{1,m}\varphi_d)^2\right)\Pi+\left((\chi-1)\,k_{1,c}A_{2,c}+k_{1,m}A_{2,m}\right)^2\left(\pi_2^2-f_{c,0}\right)$$

$$\Pi=(\pi_1-\pi_2 f_{d,0})$$

$$H_2=\Upsilon\,(\chi-1)^2\,k^2_{1,c}A^2_{2,c}+(\chi-1)\,A_{2,c}\,(k_{1,c}-1)$$
$$+0.5\left[\left((\chi-1)\,k_{1,c}A_{1,c}+k_{1,m}A_{1,m}\right)^2\varphi_c^2+\left((k_{1,m}A^d_{1,m}\varphi_d)^2+\gamma^2\right)\pi_2\right]$$

Once all the coefficient are verified, we can derive easily the conditional mean of equity premium from the fact that risk premium of any asset is negatively related with both the conditional covariance between the return and intertemporal marginal rate of substitution and variance of the return.

Then innovation to the return $r_{m,t+1}$ is

$$\begin{aligned} r_{m,t+1} - E_t\left(r_{m,t+1}\right) = {} & \varphi_d \sqrt{f_{d,t}} \epsilon_{d,t+1} + k_{1,m} A^c_{1,m} \varphi_c \sqrt{f_{c,t}} e_{c,t+1} + k_{1,m} A^d_{1,m} \varphi_d \sqrt{f_{d,t}} e_{d,t+1} \\ & + k_{1,m} A_{2,m} \left(\Gamma_1 f_{c,t} + \Gamma_2\right) u_{t+1} \end{aligned}$$

and its conditional variance $Var_t\left(r_{m,t+1}\right)$ is

$$Var_t\left(r_{m,t+1}\right) = \left(1 + \left(k_{1,m} A^d_{1,m} \varphi_d\right)^2\right) f_{d,t} + \left(k_{1,m} A^c_{1,m} \varphi_c\right)^2 f_{c,t} + \left(k_{1,m} A_{2,m} \left(\Gamma_1 f_{c,t} + \Gamma_2\right)\right)^2$$

and the risk premium for $r_{m,t+1}$ is equal to,

$$\begin{aligned} E_t[r_{m,t+1} - r_{f,t}] &= -Cov_t[m_{t+1} - E_t(m_{t+1}), r_{m,t+1} - E_t(r_{m,t+1})] - \frac{1}{2} Var_t(r_{m,t+1}) \\ &= \Lambda_{m,e} k_{1,m} A^c_{1,m} \varphi_c f_{c,t} + \Lambda_{m,u}(t) k_{1,m} A_{2,m} \left(\Gamma_1 f_{c,t} + \Gamma_2\right) - \frac{1}{2} Var_t(r_{m,t+1}) \end{aligned}$$

Now we assume the long-run component of dividend is strongly related with the long-run component of consumption following Bansal and Yaron (2004).

$$\eta_{d,t} = \phi \eta_{c,t}$$

Though our model does not assume such relation, this assumption is not bad for the illustration purpose because actual data seems to share similar trend. Then we can reduce the number of the state variable and verify undetermined coefficients as

following:

$$A_{0,m} = \frac{\left[\chi \log \delta - \gamma\mu_c + \mu_d + (\chi - 1)\left(k_{0,c} + k_{1,c}A_{0,c} - A_{0,c}\right) + k_{0,m} + 0.5H_3\right]}{1 - k_{1,m}} \tag{B.3}$$

$$A_{1,m} = \left(\frac{\phi - \psi^{-1}}{1 - k_{1,m}\rho_c}\right) \tag{B.4}$$

$$\begin{aligned} A_{2,m} = {} & \frac{1}{2\Theta k_{1,m}^2} \\ & \Bigg((1 - k_{1,m}) + 2(1 - \chi) k_{1,c}A_{2,c}k_{1,m}\Theta \\ & - \sqrt{\left[(1 - k_{1,m}) + 2(\chi - 1) k_{1,c}A_{2,c}k_{1,m}\Theta\right]^2 - 4\Theta k_{1,m}^2 \mathbb{C}} \Bigg) \end{aligned} \tag{B.5}$$

where $\Theta = \Gamma_1\left(\Gamma_1 f_{0,c} + \Gamma_2\right)$ and

$$\begin{aligned} \mathbb{C} = {} & (\chi - 1) A_{2,c}\left(k_{1,c} - 1\right) + 0.5\left(\gamma^2 + \pi_2 + \left((\chi - 1) k_{1,c}A_{1,c} + k_{1,m}A_{1,m}\right)^2\right) \\ & + \Theta(\chi - 1)^2 k_{1,c}^2 A_{2,c}^2 \\ H_3 = {} & \Pi + \left((\chi - 1) k_{1,c}A_{2,c} + k_{1,m}A_{2,m}\right)^2\left(\pi_2^2 - f_{c,0}\right) \end{aligned}$$

Then innovation to the return $r_{m,t+1}$ is

$$r_{m,t+1} - E_t\left(r_{m,t+1}\right) = \varphi_d\sqrt{f_{d,t}}\epsilon_{d,t+1} + k_{1,m}A_{1,m}\varphi_c\sqrt{f_{c,t}}e_{c,t+1} + k_{1,m}A_{2,m}\left(\Gamma_1 f_{c,t} + \Gamma_2\right)u_{t+1}$$

and its conditional variance $Var_t\left(r_{m,t+1}\right)$ is

$$Var_t\left(r_{m,t+1}\right) = \varphi_d^2 f_{d,t} + k_{1,m}^2 A_{1,m}^2 \varphi_c^2 f_{c,t} + k_{1,m}^2 A_{2,m}^2\left(\Gamma_1 f_{c,t} + \Gamma_2\right)^2$$

and the risk premium for $r_{m,t+1}$ is equal to,

$$\begin{aligned} E_t[r_{m,t+1} - r_{f,t}] &= -Cov_t[m_{t+1} - E_t(m_{t+1}), r_{m,t+1} - E_t(r_{m,t+1})] - \frac{1}{2}Var_t(r_{m,t+1}) \\ &= \Lambda_{m,e}k_{1,m}A_{1,m}\varphi_c f_{c,t} + \Lambda_{m,u}(t)k_{1,m}A_{2,m}\left(\Gamma_1 f_{c,t} + \Gamma_2\right) - \frac{1}{2}Var_t(r_{m,t+1}) \end{aligned}$$

APPENDIX C

SOLUTION TO THE RISK-FREE RATE

The risk-free rate can be derived from the Euler equation.

$$\begin{aligned} r_{f,t} &= -\log\left[E_t \exp(m_{t+1})\right] \\ &= -\log\delta + \frac{1}{\psi}E_t g_{c,t+1} + \frac{1-\chi}{\chi}E_t\left[r_{w,t+1} - r_{f,t}\right] - \frac{1}{2\chi}Var_t(m_{t+1}) \end{aligned}$$

Here we have three terms to elaborate. First, the conditional mean of consumption growth $E_t g_{c,t+1}$ is simply $\mu_c + \eta_{c,t}$. Secondly, the conditional mean of excess return on the wealth $E_t\left[r_{w,t+1} - r_{f,t}\right]$ is can be solved by the conditional covariance between $r_{w,t+1}$ and m_{t+1} and conditional variance of $r_{w,t+1}$. Using the fact that $r_{w,t+1} - E_t r_{w,t+1} = \sqrt{f_{c,t}}\epsilon_{t+1} + \frac{\Lambda_{m,e}}{1-\chi}\sqrt{f_{c,t}}e_{t+1} + \frac{\Lambda_{m,u}(t)}{1-\chi}u_{t+1}$ we have

$$\begin{aligned} E_t\left[r_{w,t+1} - r_{f,t}\right] &= -Cov_t\left[m_{t+1} - E_t(m_{t+1}), r_{w,t+1} - E_t(r_{w,t+1})\right] - \frac{1}{2}Var_t(r_{w,t+1}) \\ &= \left[\gamma + \frac{\Lambda_{m,e}^2}{1-\chi} - \frac{1}{2}\left(\frac{\Lambda_{m,e}^2}{(1-\chi)^2} + 1\right)\right] f_{c,t} + \frac{\Lambda_{m,u}^2(t)}{(1-\chi)^2}\left(\frac{1}{2} - \chi\right) \end{aligned}$$

Lastly, the conditional variance of IMRS is derived from (A.6)

$$Var_t(m_{t+1}) = \Lambda_{m,\epsilon}^2 f_{c,t} + \Lambda_{m,e}^2 f_{c,t} + \Lambda_{m,u}^2(t)$$

APPENDIX D

CONVENTIONAL DENSITY-BASED NONLINEAR FILTERING ALGORITHM

If we apply the nonlinear filtering algorithm from Tanizaki (1996), we can get the following.

Prediction Step:

$$\begin{aligned} p(w_t|\mathcal{F}_{t-1}) &= \int p(w_t, w_{t-1}|\mathcal{F}_{t-1})\, dw_{t-1} \\ &= \int p(w_t|w_{t-1}) p(w_{t-1}|\mathcal{F}_{t-1})\, dw_{t-t} \\ p(x_{c,t}, x_{d,t}|\mathcal{F}_{t-1}) &= \int p(x_{c,t}, x_{d,t}, w_{t-1}|\mathcal{F}_{t-1})\, dw_{t-1} \\ &= \int p(x_{c,t}, x_{d,t}|w_{t-1}) p(w_{t-1}|\mathcal{F}_{t-1})\, dw_{t-t} \end{aligned}$$

Updating Step:

$$\begin{aligned} p(w_t|\mathcal{F}_t) &= p(w_t, g_{c,t}, g_{d,t}|\mathcal{F}_{t-1}) / p(g_{c,t}, g_{d,t}|\mathcal{F}_{t-1}) \\ &= p(g_{c,t}, g_{d,t}|w_t) p(w_t|\mathcal{F}_{t-1}) / p(g_{c,t}, g_{d,t}|\mathcal{F}_{t-1}) \\ p(x_{c,t}, x_{d,t}|\mathcal{F}_t) &= p(x_{c,t}, x_{d,t}, g_{c,t}, g_{d,t}|\mathcal{F}_{t-1}) / p(g_{c,t}, g_{d,t}|\mathcal{F}_{t-1}) \\ &= p(g_{c,t}, g_{d,t}|x_{1t}, x_{2t}) p(x_{1t}, x_{2t}|\mathcal{F}_{t-1}) / p(g_{c,t}, g_{d,t}|\mathcal{F}_{t-1}) \end{aligned}$$

where

$$\begin{aligned} p(g_{c,t}, g_{d,t}|w_t) &= \iint p(g_{c,t}, g_{d,t}, x_{c,t}, x_{d,t}|w_t)\, dx_{c,t}\, dx_{d,t} \\ &= \iint p(g_{c,t}, g_{d,t}|x_{c,t}, x_{d,t}) p(x_{c,t}, x_{d,t}|w_t)\, dx_{c,t}\, dx_{d,t} \end{aligned}$$

APPENDIX E

GIBBS SAMPLING ALGORITHM

First, the marginal density for w_t is

$$\begin{aligned} & p(w_t|X_c, X_d, W_{\setminus t}, Y, \Psi) \\ & \quad \propto p(w_t, x_{c,t}, x_{d,t}, w_{t+1}|w_{t-1}, \Psi) \\ & \quad \propto p(x_{c,t}, x_{d,t}|w_t, w_{t+1}, w_{t-1}, \Psi)p(w_{t+1}, w_t|w_{t-1}) \\ & \quad = p(x_{c,t}|w_t, \lambda_c, \sigma_c^2)p(x_{d,t}|w_t, \lambda_d, \sigma_d^2)p(w_{t+1}|w_t)p(w_t|w_{t-1}) \end{aligned}$$

We know that the marginal density of $x_{j,t}|w_t$ follows a normal distribution $N\left(\lambda_j w_t, \sigma_j^2\right)$. And given one period past values, the density of w_{t+1} and w_t are known. Combining the four normal densities, we have a normal distribution $\mathbb{N}(BA^{-1}, A^{-1})$ where $A = \frac{\lambda_c^2}{\sigma_c^2} + \frac{\lambda_d^2}{\sigma_d^2} + 2$ and $B = \frac{\lambda_c x_{c,t}}{\sigma_c^2} + \frac{\lambda_d x_{d,t}}{\sigma_d^2} + w_{t+1} + w_{t-1}$.

Secondly, the marginal density for $x_{c,t}$ is

$$\begin{aligned} p(x_{c,t}|X_{\setminus c,t}, W, Y, \Psi) &= p(x_{c,t}|x_{d,t}, w_t, y_t, \Psi) \\ &\propto p(x_t, y_t|w_t, \Psi) \\ &\propto p(y_t|x_t, w_t, \Psi)p(x_t|w_t, \Psi) \\ &\propto p(y_t|x_t, \theta, \rho)p(x_{c,t}|w_t, \lambda_c, \sigma_c^2) \end{aligned}$$

where $p(y_t|x_t, \theta, \rho) = \frac{1}{2\pi}|S_t \Sigma S_t'|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} y_t' S_t^{-1} \Sigma^{-1} S_t^{-1} y_t\right)$ and $p(x_{c,t}|w_t, \lambda_c, \sigma_c^2) = \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left(-\frac{(x_{c,t} - \lambda_c w_t)^2}{2\sigma_c^2}\right)$. Similarly, we could get the marginal density for $x_{d,t}$. Here we denote S_t, Σ, and y_t as following:

$$S_t = \begin{pmatrix} \sqrt{f(x_{c,t})} & 0 \\ 0 & \sqrt{f(x_{d,t})} \end{pmatrix} \quad \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \quad y_t = \begin{pmatrix} y_{c,t} \\ y_{d,t} \end{pmatrix}$$

Thirdly, the marginal density for θ_j is

$$p(\theta_j|X,W,Y,\Psi_{\backslash\theta_j}) \propto p(Y|X,\theta,\rho)p(\theta_j)$$
$$= \prod_{t=1}^{n} \frac{1}{2\pi}|S_t\Sigma S_t'|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}y_t'S_t^{-1}\Sigma^{-1}S_t^{-1}y_t\right) p\left(\theta_j\right)$$

and for ρ,

$$p(\rho|X,W,Y,\Psi_{\backslash\rho}) \propto p(Y|X,\theta,\rho)p\left(\rho\right)$$
$$= \prod_{t=1}^{n} \frac{1}{2\pi}|S_t\Sigma S_t'|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}y_t'S_t^{-1}\Sigma^{-1}S_t^{-1}y_t\right) p\left(\rho\right)$$

Lastly, the marginal density for Ψ_2 is

$$p(\Psi_{2k,j}|X,W,Y,\Psi_{\backslash 2k,j}) \propto p(X_j|W,\Psi_{2,j})p(\Psi_{2k,j})$$
$$= \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(x_{j,t}-\lambda_j w_t)^2}{2\sigma_j^2}\right) p(\Psi_{2k,j})$$

where $\Psi_{21,j} = \lambda_j$ and $\Psi_{22,j} = \sigma_j^2$.

VITA

Hyoung Il Lee received his Bachelor of Arts degree in economics from Seoul National University in 1993. He received his Ph.D. in Economics from Texas A&M University in August of 2008. His fields of specialization are Econometrics and Macroeconomics. Hyoung Il Lee can be reached at Ministry of Strategy and Finance, 88 Kwanmun-ro, Kwacheon-si, Kyunggi-do, 427-725, Republic of Korea, or by email at leehi@mosf.go.kr.